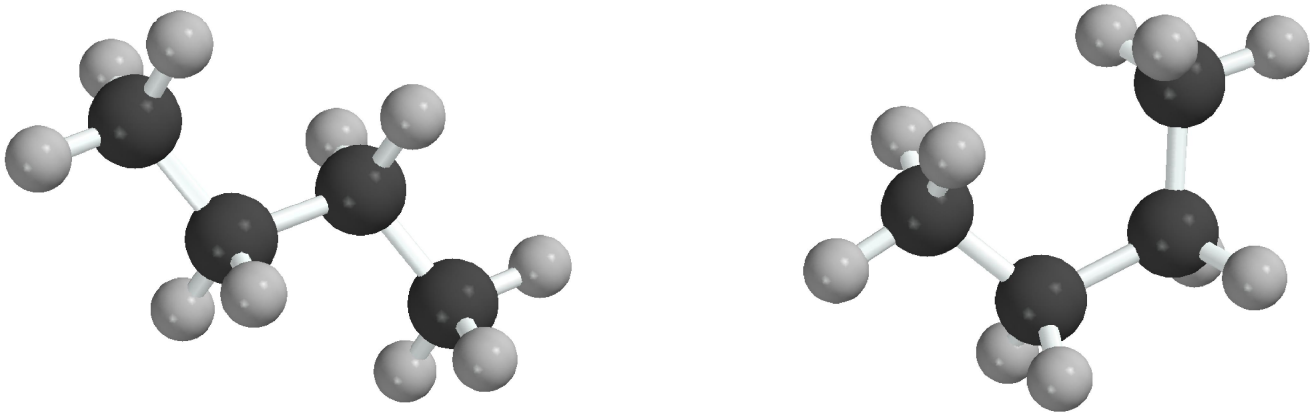
M.Sc. Bioinformatics

UNIT-I

Molecular Modeling Part I
Molecular Mechanics and Conformational Analysis



ORG I Lab

William Kelly

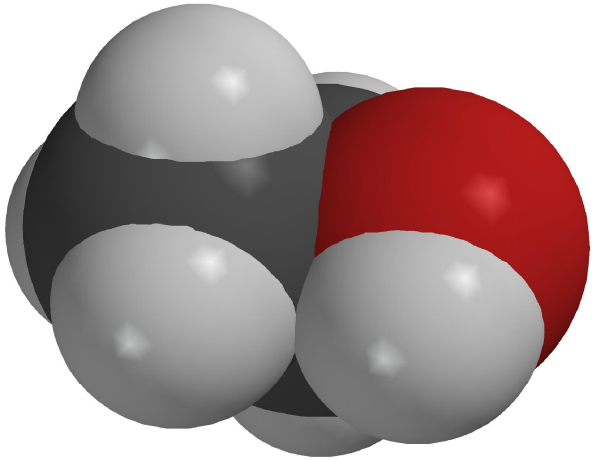
Molecular Modeling

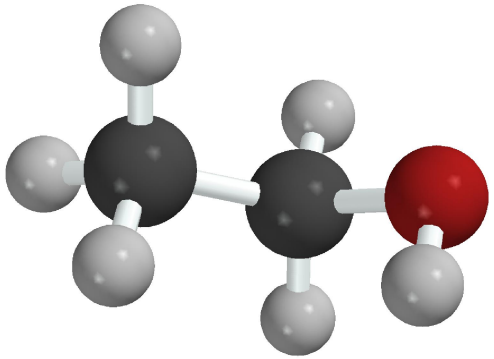
Molecular modeling in the broadest sense is the use of:

a) Physical representations: I.e. Plastic Molecular Models

b) Graphical representations:

Space Filling Model

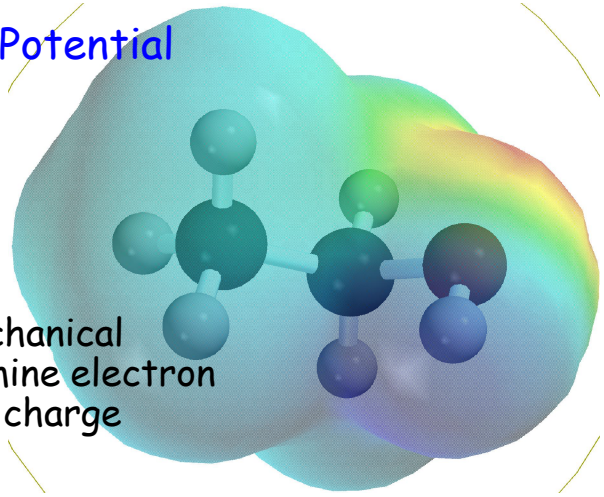




Ball and Spoke Model

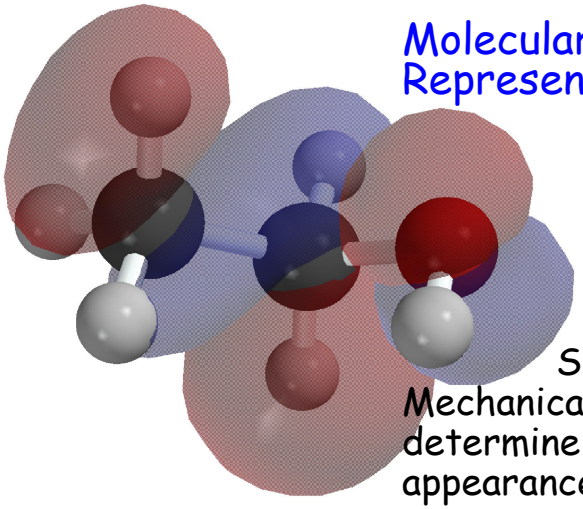
c) Mathematical representations

Electrostatic Potential Map,



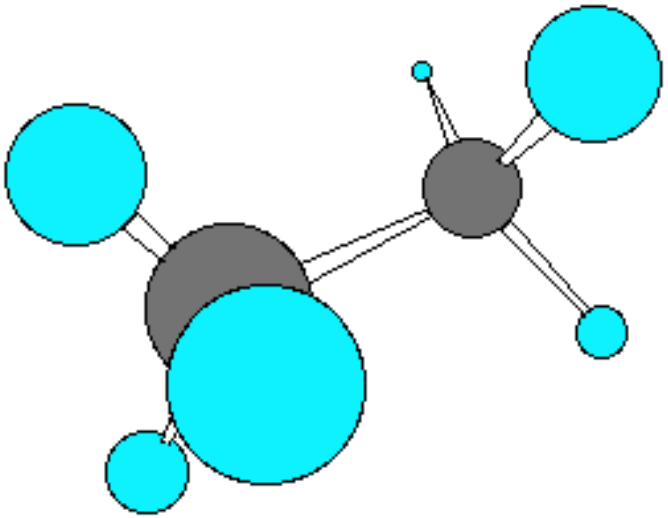
Solve Quantum Mechanical equations to determine electron position and atomic charge

Molecular Orbital Representation



Solve Quantum Mechanical equations to determine molecular orbital appearance

Molecular Mechanics Simulation of Dynamic Motion in Ethane



Combination of Mathematical and Graphic Model

Animation obtained from the solution of molecular mechanics equations of atomic motion for atoms of ethane. Dynamic motion followed for 1 picosec (10^{-9} sec). Each frame represents a 3 femtosec (3×10^{-12} sec) window.

With the development of fast, easy to use computers within the last 15 years computer modeling of molecules has become an important tool for the practicing chemist. There are many approaches to this subject, and a gradation of levels of theory in the understanding of molecules through computer modeling.

COMPUATIONAL METHOD	MOLECULE SIZE
MOLECULAR MECHANICS	1,000's atoms
SEMI-EMPIRICAL QUANTUM MECHANICS	100's atoms
ab initio QUANTUM MECHANICS	up to 100 atoms
CORRELATED QUANTUM MECHANICS	up to 40 atoms
CORRELATED, RELATIVISTIC QUANTUM MECHANICS	up to 10 atoms

Increasing Accuracy ↓ ↑ **Decreasing Time**

The subject of this experiment is a technique called ***Force Field Molecular Modeling or Molecular Mechanics***.

Molecular Mechanics Background

The "mechanical" molecular model was developed out of a need to describe molecular structures and properties in as practical a manner as possible.

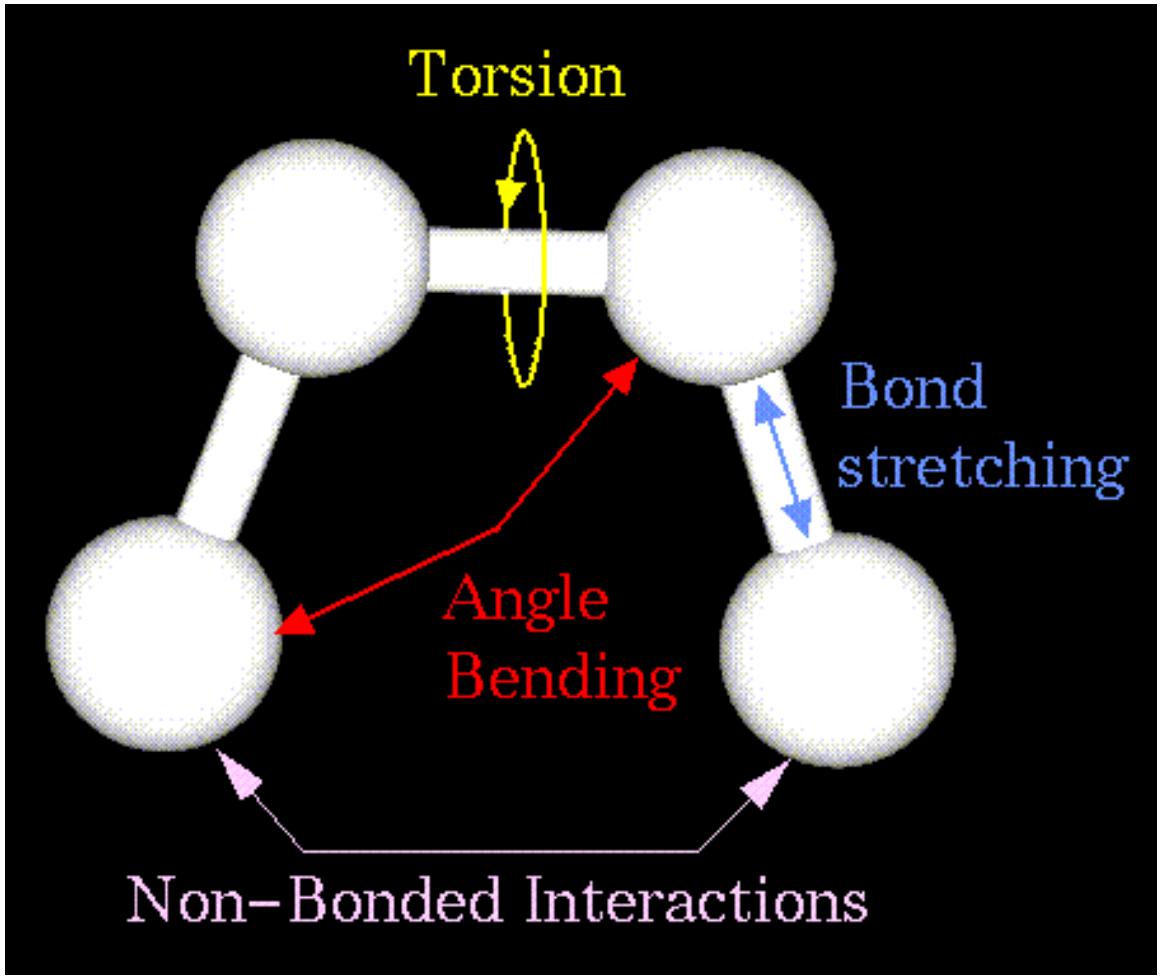
- The great computational speed of molecular mechanics allows for its use in molecules containing thousands of atoms.

Molecular mechanics methods are based on the following principles:

- Nuclei and electrons are lumped into atom-like particles.
- Atom-like particles are spherical (radii obtained from measurements or theory) and have a net charge (obtained from theory).
- Interactions are based on springs and classical potentials.
- Interactions must be preassigned to specific sets of atoms.
- Interactions determine the **spatial distribution** of atom-like particles and their **energies**.

The Anatomy of a Molecular Mechanics Force-Field

The mechanical molecular model considers atoms as spheres and bonds as springs. The mathematics of spring deformation can be used to describe the ability of bonds to stretch, bend, and twist:



Non-bonded atoms (greater than two bonds apart) interact through van der Waals attraction, steric repulsion, and electrostatic attraction/repulsion. These properties are easiest to describe mathematically when atoms are considered as spheres of characteristic radii.

The Anatomy of a Molecular Mechanics Force-Field

The object of molecular mechanics is to predict the **energy** associated with a given conformation of a molecule. However, molecular mechanics energies have no meaning as absolute quantities. Only differences in energy between two or more conformations have meaning. A simple molecular mechanics energy equation is given by:

Total Energy =Stretching Energy +Bending Energy +Torsion Energy +Non-Bonded Interaction Energy*

Within the molecular framework, the "total energy" of a molecule is described in terms of a sum of contributions arising from ** **ALL DEVIATIONS**** from "ideal" bond distances (stretch contributions), bond angles (bend contributions) and dihedral angles (torsion contributions) summarized by

$$E^{Total} = \sum_{i}^{bonds} E_i^{stretch} + \sum_{i}^{bondangles} E_i^{bend} + \sum_{i}^{dihedralangles} E_i^{torsion} + \sum_{ij}^{atompairs} E^{vanderwaals} + \sum_{ij}^{atompairs} E^{electrostatic}$$

Covalent Interactions Non- covalent Interactions

The Anatomy of a Molecular Mechanics Force-Field

These equations together with the data (parameters) required to describe the behavior of different kinds of atoms and bonds, is called a "**FORCE FIELD**".

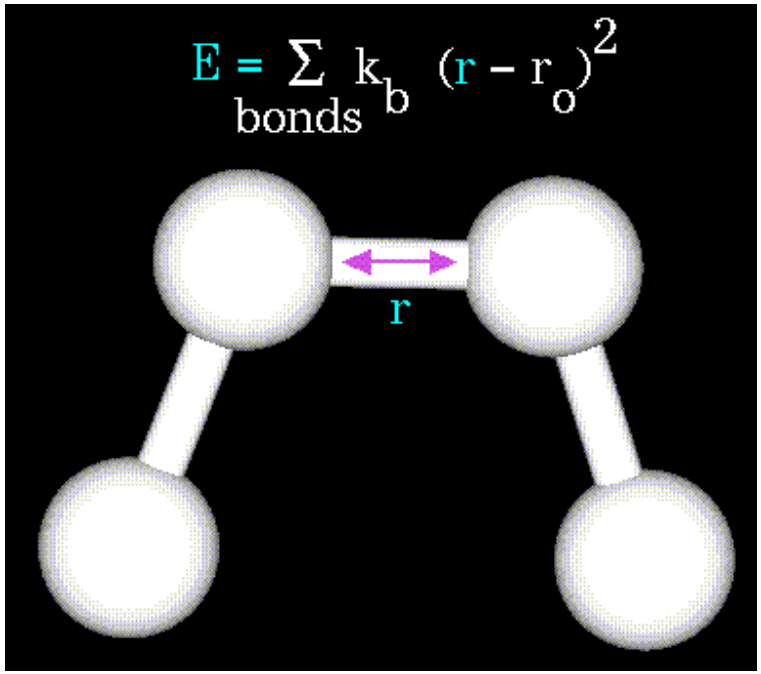
The molecular mechanics "FORCE FIELD" relates the motions, and energies of motions of atoms within the molecule. The force field is used to govern how the parts of a molecule relate to each other, that is, how each atom or group of atoms *is affected by its environment*, and how these forces contribute to the structure of the molecule.

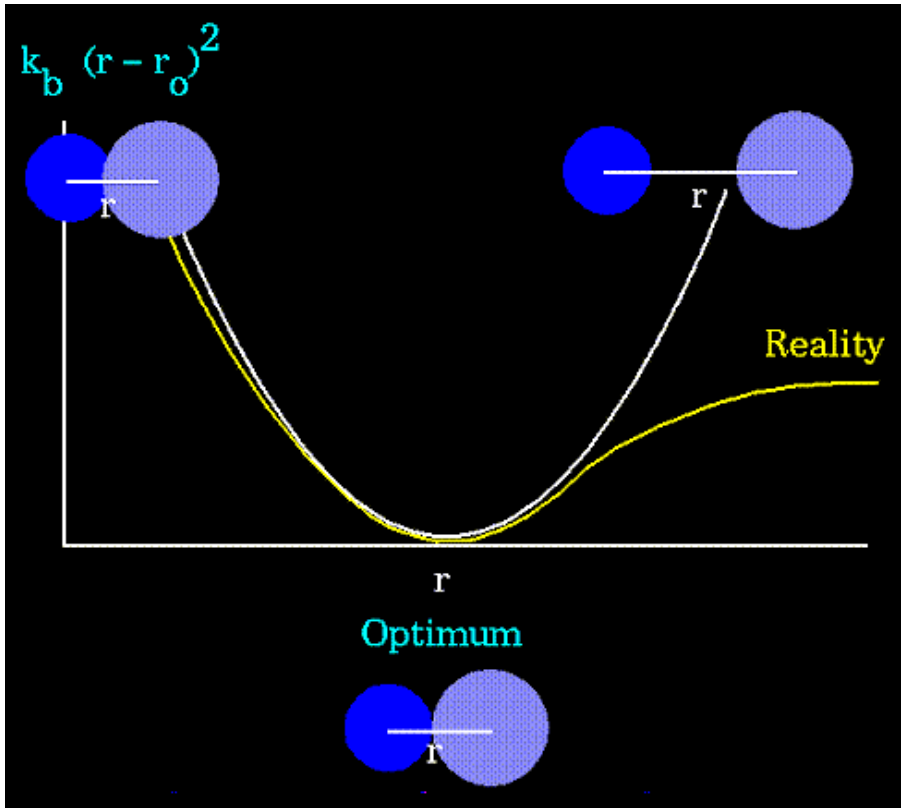
Many different kinds of force-fields have been developed over the years. Some include additional energy terms that describe other kinds of deformations. Some force-fields account for coupling between bending and stretching in adjacent bonds in order to improve the accuracy of the mechanical model.

The mathematical form of the energy terms varies from force-field to force-field. The more common forms will be described

STRETCHING ENERGY

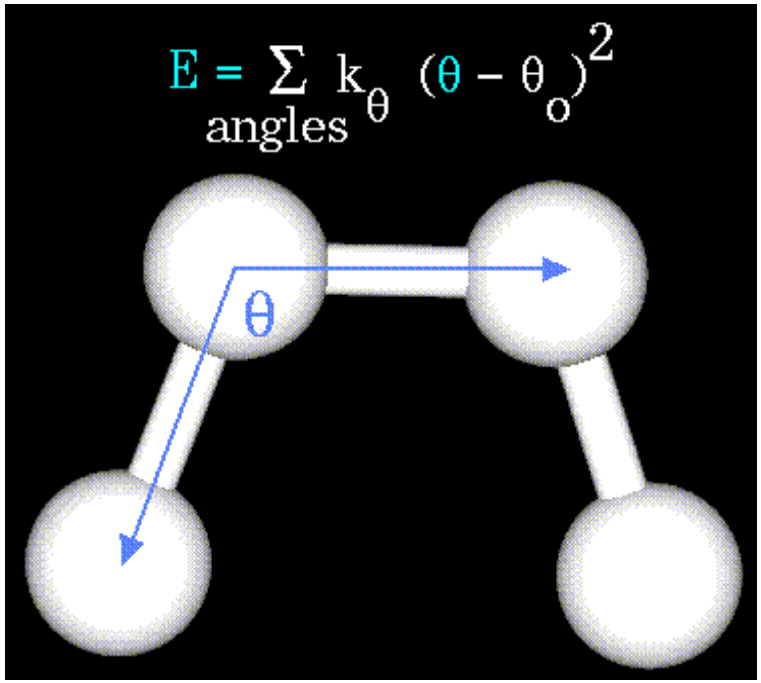
•The stretching energy equation is based on Hooke's law. The "k_b" parameter controls the stiffness of the bond spring, while "r_o" defines its equilibrium length. Unique "k_b" and "ro" parameters are assigned to each pair of bonded atoms based on their types (e.g. C-C, C-H, O-C, etc.). This equation estimates the energy associated with vibration about the equilibrium bond length. This is the equation of a parabola, as can be seen in the following plot

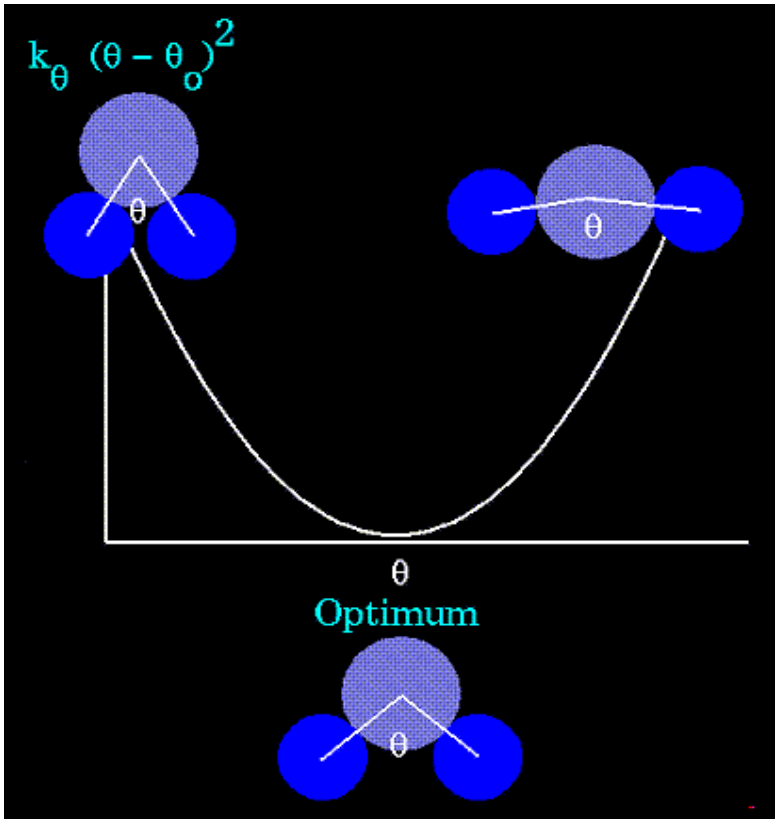




BENDING ENERGY

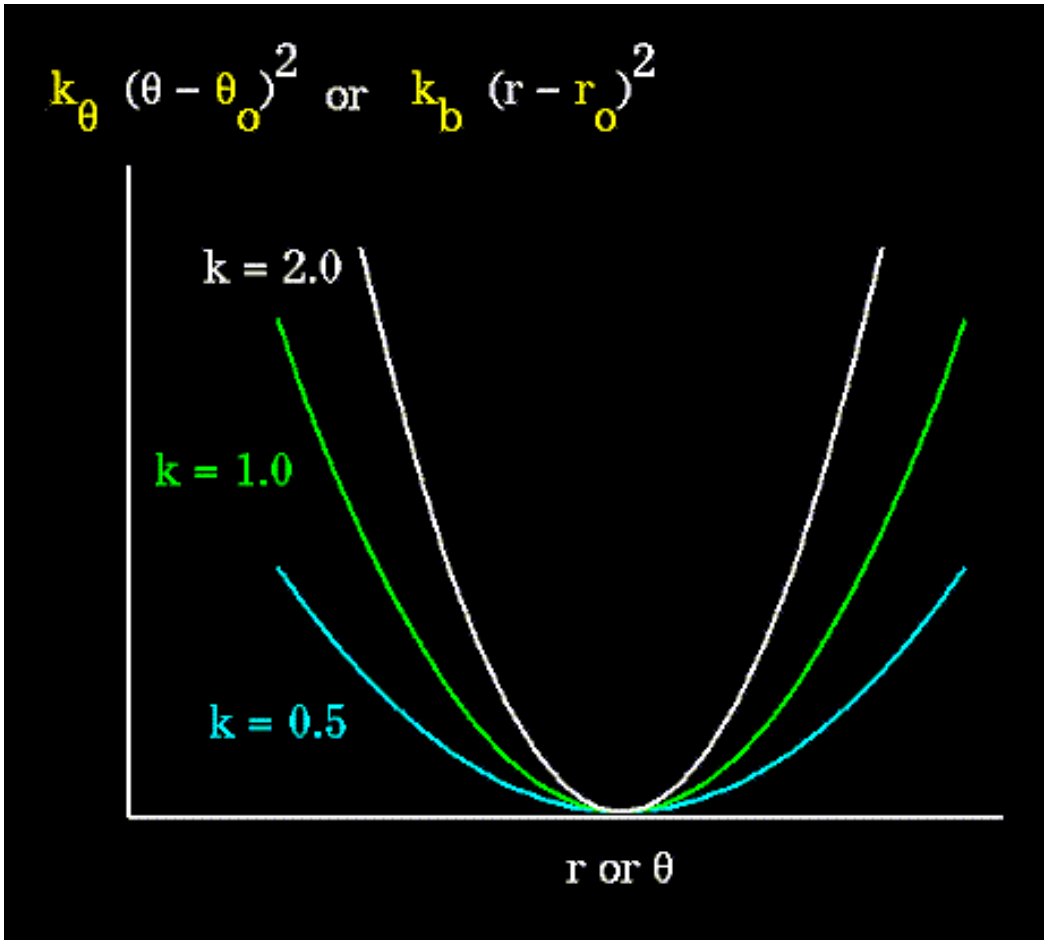
•The bending energy equation is also based on Hooke's law. The "k_θ" parameter controls the stiffness of the angle spring, while "θ_o" defines its equilibrium angle. This equation estimates the energy associated with vibration about the equilibrium bond angle





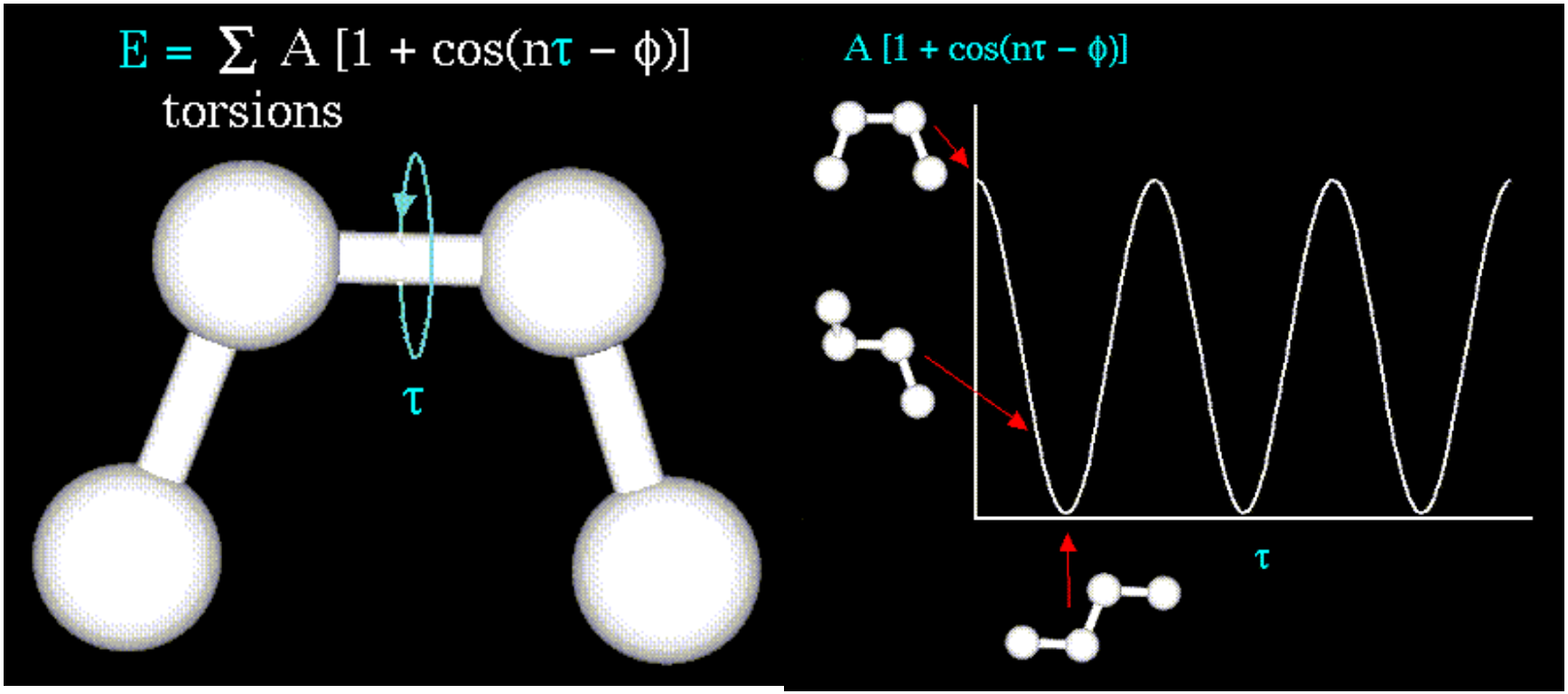
UNIQUE STRETCHING AND BENDING ENERGY

•Unique parameters for angle bending are assigned to each bonded triplet of atoms based on their types (e.g. *C-C-C*, *C-O-C*, *C-C-H*, etc.). The effect of the "k_b" and "k_θ" parameters is to broaden or steepen the slope of the parabola. The larger the value of "k", the more energy is required to deform an angle (or bond) from its equilibrium value. Shallow potentials are achieved for "k" values between 0.0 and 1.0. The Hookeian potential is shown in the following plot for three values of "k



TORSIONAL ENERGY

The torsion energy is modeled by a simple periodic function

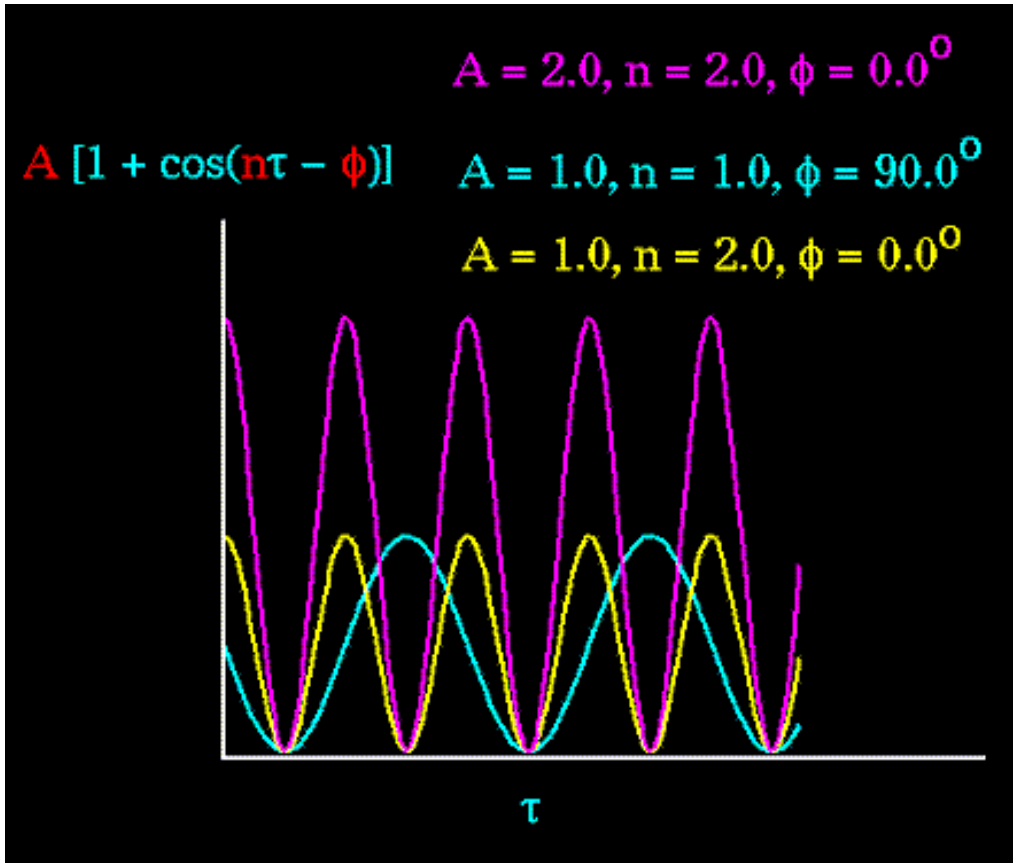


Torsional energy varies during rotation about *C-C*, *C-N* and *C-O* single bonds. The maximum values occur at t=0° and represent "eclipsing" interactions between atoms separated by three sigma bonds.

TORSIONAL ENERGY

The torsion energy is modeled by a simple periodic function

The "A" parameter controls the amplitude of the curve, the n parameter controls its periodicity, and "phi" shifts the entire curve along the rotation angle axis (tau). The parameters are determined from curve fitting. Unique parameters for torsional rotation are assigned to each bonded quartet of atoms based on their types (e.g. C-C-C-C, C-O-C-N, H-C-C-H, etc.). Torsion potentials with three combinations of "A", "n", and "phi" are shown



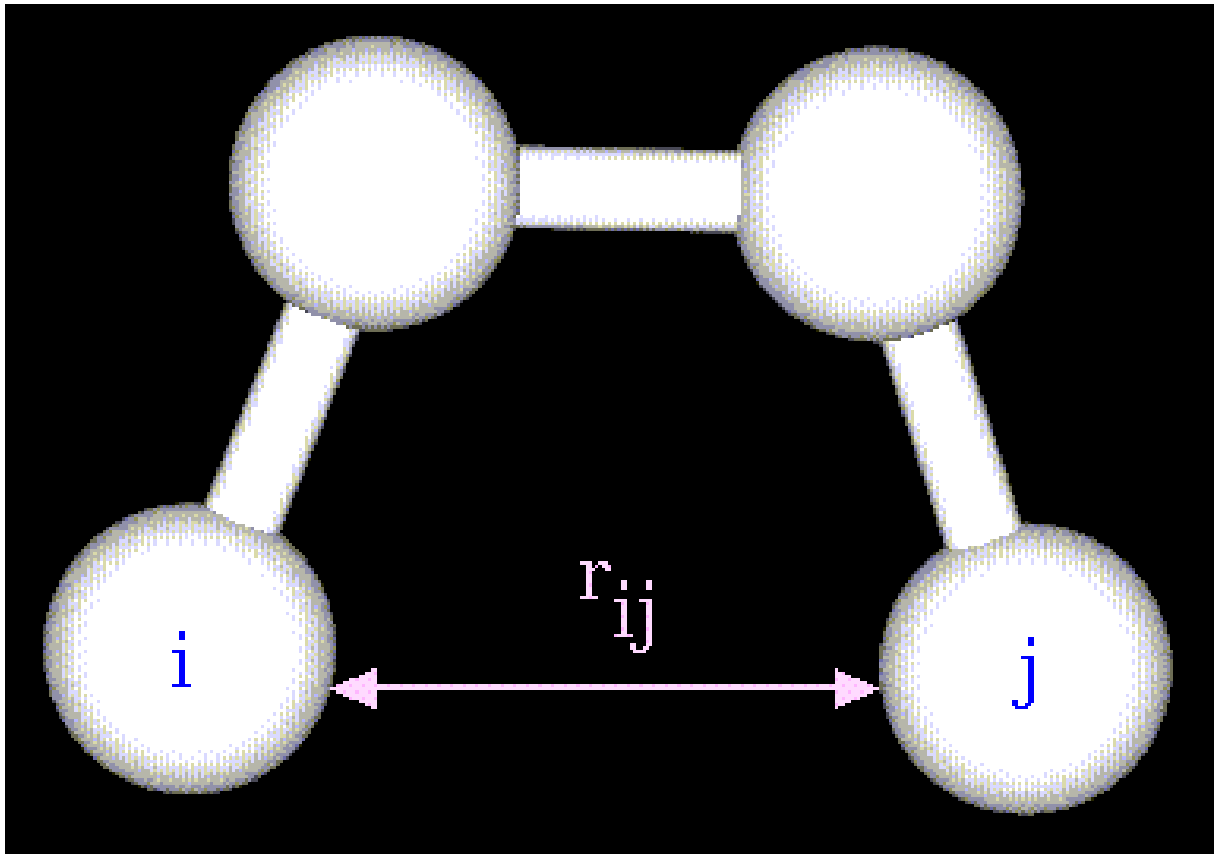
NON-COVALENT (NON-BONDED) TWO ATOM INTERACIONS

The non-bonded energy represents the pair-wise sum of the energies of all possible interacting non-bonded atoms i and j

The non-bonded energy accounts for van der Waals attraction, repulsion and electrostatic interactions.

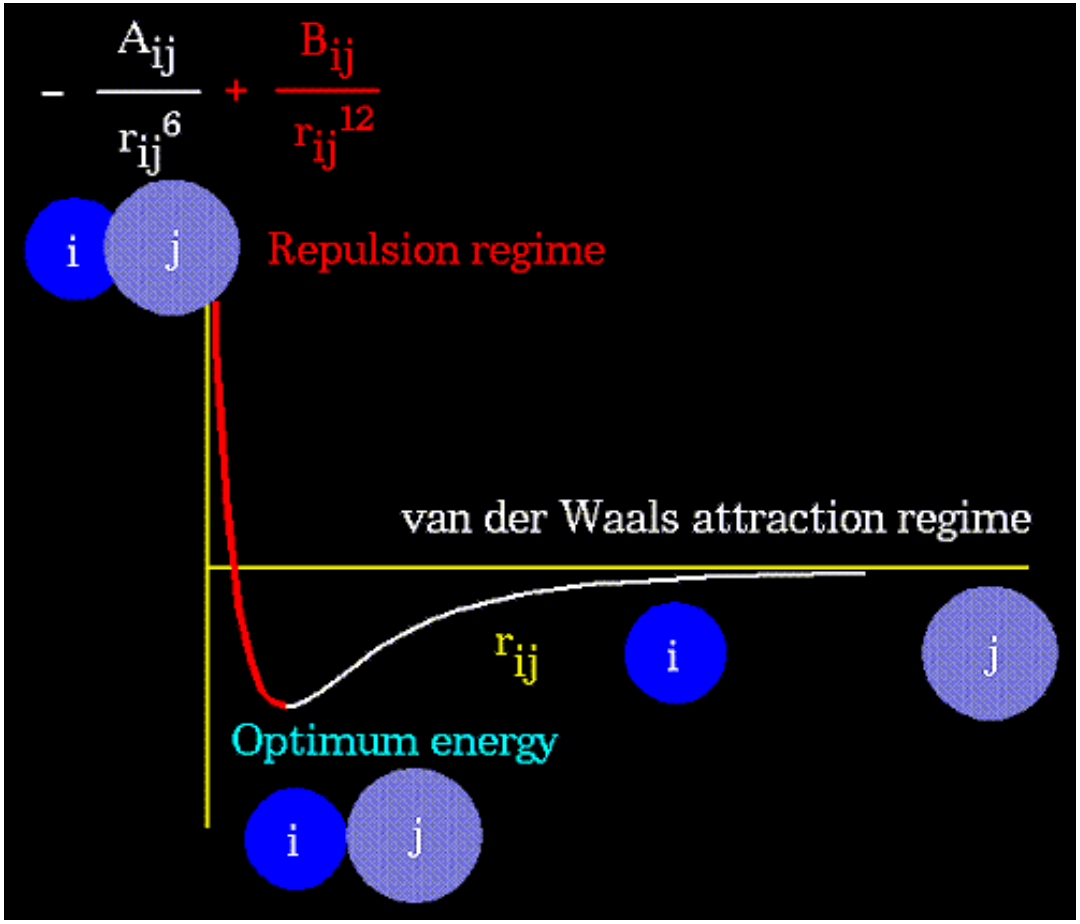
$$E = \sum_i \sum_j \frac{-A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + \sum_i \sum_j \frac{q_i q_j}{r_{ij}}$$

van der Waals term — Electrostatic term



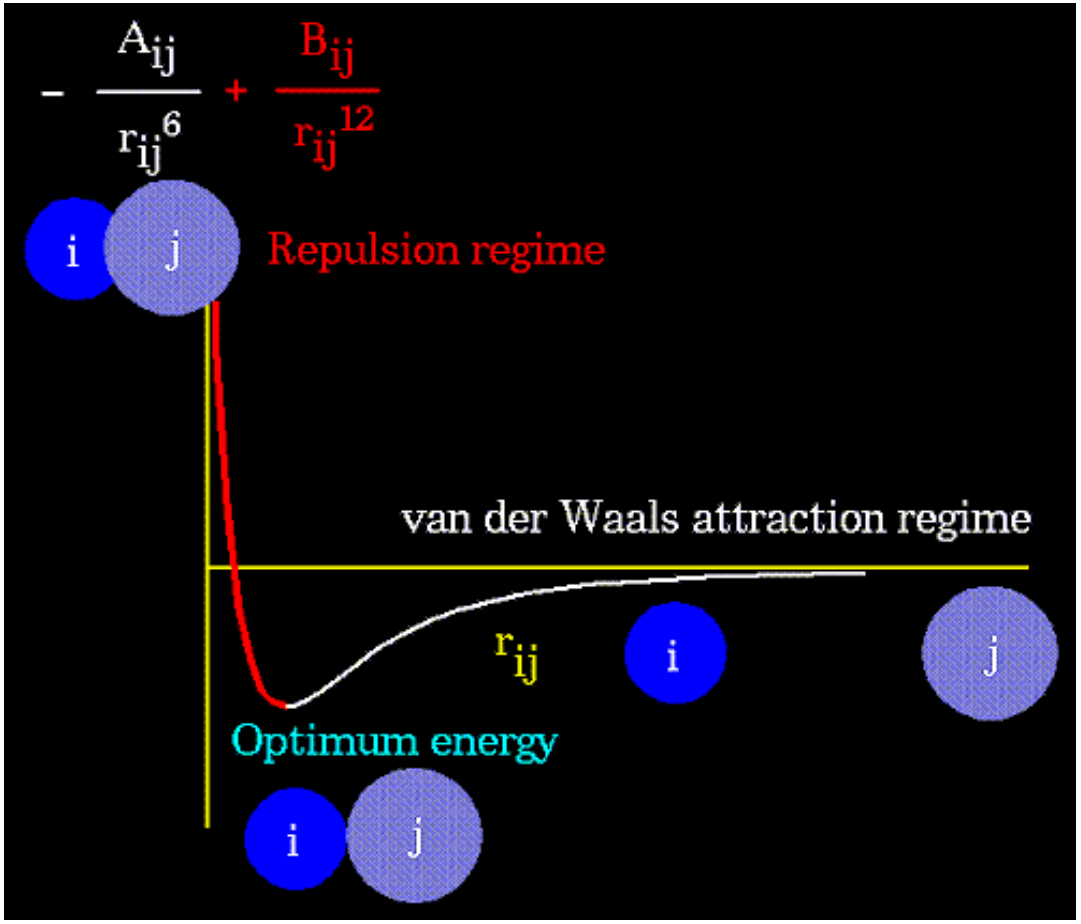
VAN DER WAALS TWO ATOM INTERACIONS

The van der Waals attraction occurs at short range, and rapidly dies off as the interacting atoms move apart by a few Angstroms. Repulsion occurs when the distance between interacting atoms becomes even slightly less than the sum of their contact radii. Repulsion is modeled by an equation that is designed to rapidly blow up at close distances ($1/r^{12}$ dependency). The energy term that describes attraction/repulsion provides for a smooth transition between these two regimes. These effects are often modeled using a 6-12 equation, as shown in the following plot



VAN DER WAALS TWO ATOM INTERACIONS

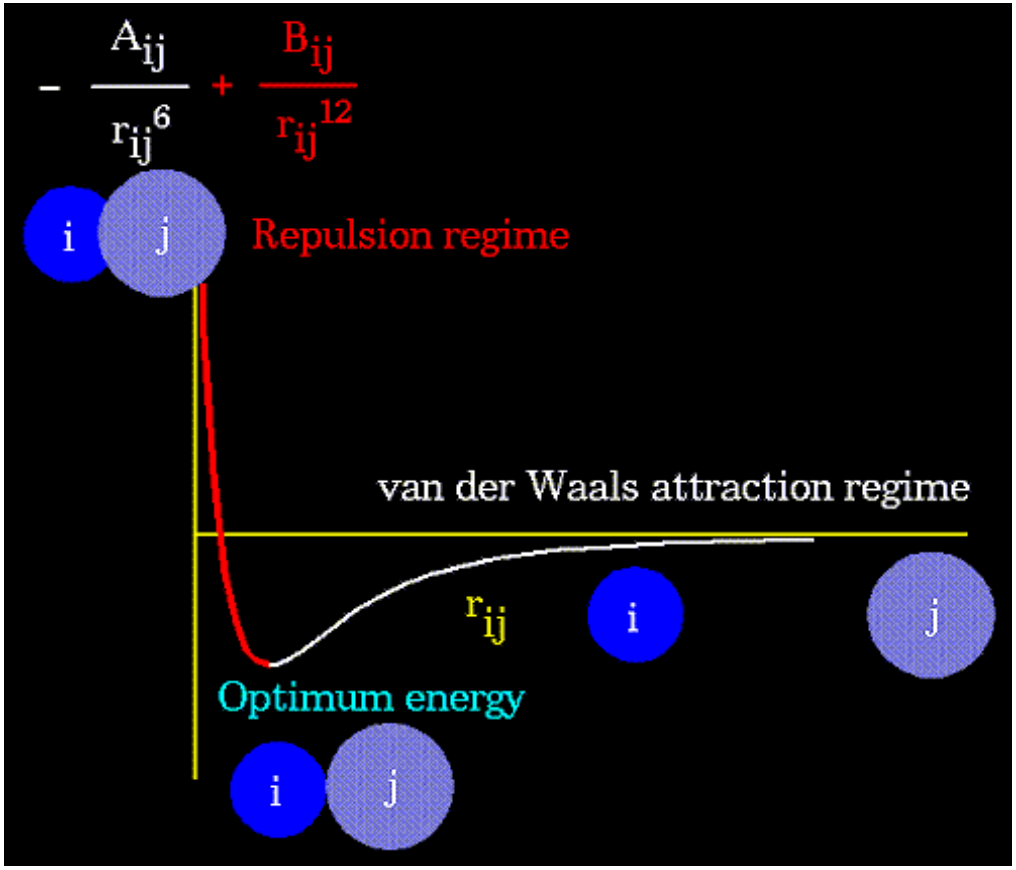
The "A" and "B" parameters control the depth and position (interatomic distance) of the potential energy well for a given pair of non-bonded interacting atoms (e.g. C:C, O:C, O:H, etc.). In effect, "A" determines the degree of "stickiness" of the van der Waals attraction and "B" determines the degree of "hardness" of the atoms (e.g marshmallow-like, billiard ball-like, etc.



The "A" parameter can be obtained from atomic polarizability measurements, or it can be calculated quantum mechanically. The "B" parameter is typically derived from crystallographic data so as to reproduce observed average contact distances between different kinds of atoms in crystals of various molecules.

VAN DER WAALS TWO ATOM INTERACIONS

TheThe "A" and "B" parameters control the depth and position (interatomic distance) of the potential energy well for a given pair of non-bonded interacting atoms (e.g. C:C, O:C, O:H, etc.). In effect, "A" determines the degree of "stickiness" of the van der Waals attraction and "B" determines the degree of "hardness" of the atoms (e.g marshmallow-like, billiard ball-like, etc.



The "A" parameter can be obtained from atomic polarizability measurements, or it can be calculated quantum mechanically. The "B" parameter is typically derived from crystallographic data so as to reproduce observed average contact distances between different kinds of atoms in crystals of various molecules.

UNIT-II
&
UNIT-III

Molecular Modeling Methods

- **Molecular modeling methods** are the theoretical methods and computational techniques used to simulate the behavior of molecules and molecular systems
- **Molecular Forcefields**
- **Conformational Search methods**
  - Energy Minimization
  - Molecular Dynamics
  - Monte Carlo simulation
  - Genetic Algorithm

Ab Initio Protein Structure Prediction

- ***Ab initio* protein structure prediction** methods build protein 3D structures from sequence based on physical principles.
- **Importance**
  - The *ab initio* methods are important even though they are computationally demanding
  - *Ab initio* methods predict protein structure based on physical models, they are indispensable complementary methods to Knowledge-based approach

    eg.

    Knowledge-based approach would fail in following conditions:
    - Structure homologues are not available
    - Possible undiscovered new fold exists

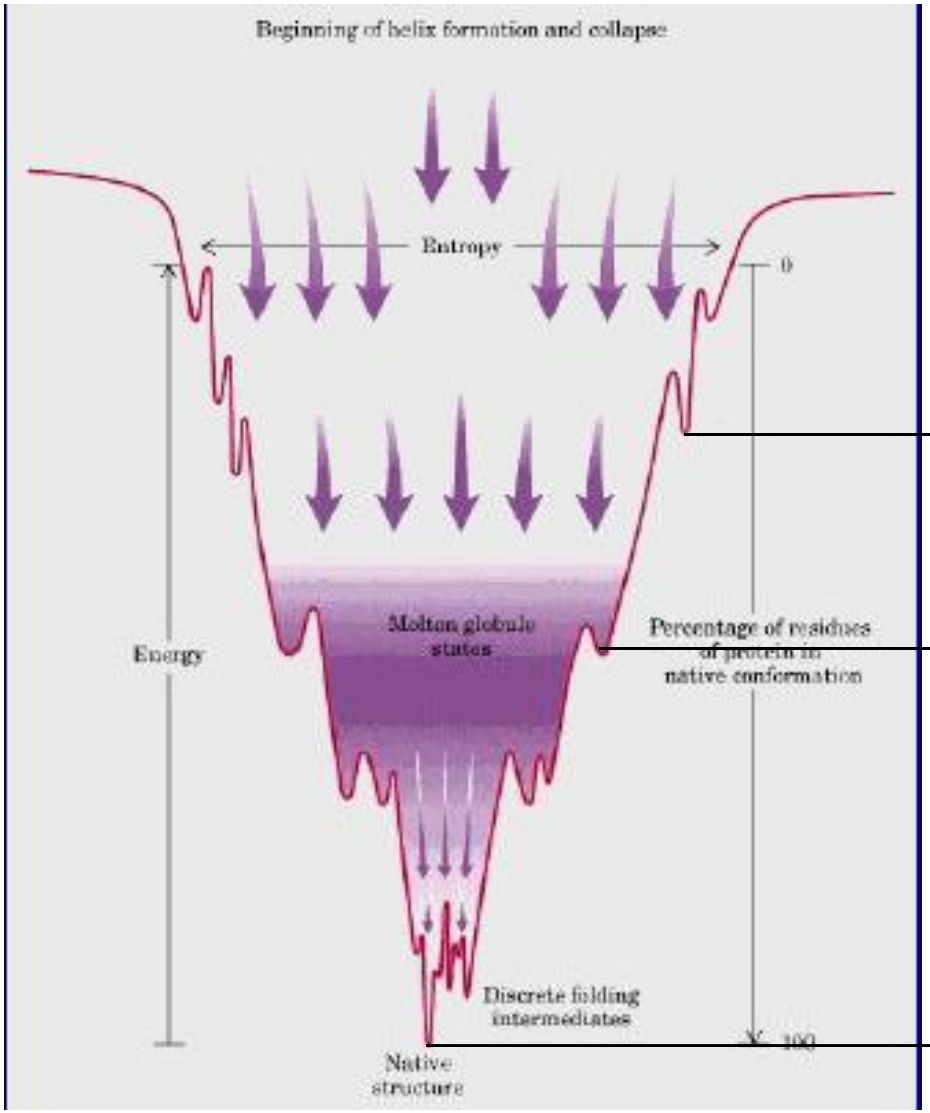
Applications of MM in *Ab Initio* PSP

- **Basic idea**

  Anfinsen's theory: Protein native structure corresponds to the state with the lowest free energy of the protein-solvent system.

- **General procedures**
  - **Potential function**
    - Evaluate the energy of protein conformation
    - Select native structure
  - **Conformational search algorithm**
    - To produce new conformations
    - Search the potential energy surface and locate the global minimum (native conformation)

Protein Folding Funnel



Local mimina

Global minimum

Native Structure

Potential Functions for PSP

- **Potential function**
  - **Physical based energy function**

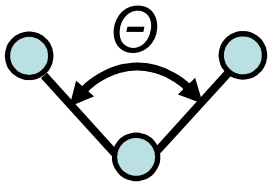
    Empirical *all-atom* forcefields: CHARMM, AMBER, ECEPP-3, GROMOS, OPLS

    Parameterization: Quantum mechanical calculations, experimental data

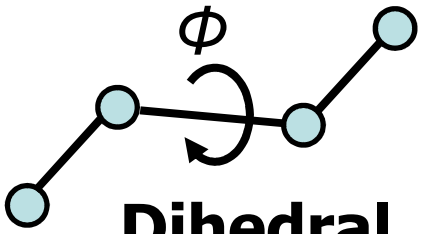
    Simplified potential: UNRES (*united residue*)
  - **Solvation energy**
    - Implicit solvation model: Generalized Born (GB) model, surface area based model
    - Explicit solvation model: TIP3P (computationally expensive)

General Form of All-atom Forcefields

Bond stretching term



Angle bending term



Dihedral term

$$V_{\text{total}} = \sum_{\text{bonds}} K_b\left(r - r_0\right)^2 + \sum_{\text{angles}} K_\theta\left(\theta - \theta_0\right)^2 + \sum_{\text{dihedrals}} K_\phi\left[1 + \cos\left(n\phi - \gamma\right)\right]$$

$$+ \sum_{\text{Hbonds}}\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + \sum_{\substack{\text{van der Waals} \\ i,j \text{ pairs}}}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + \sum_{\substack{\text{electrostatic} \\ i,j \text{ pairs}}} \frac{q_i q_j}{\varepsilon r_{ij}}$$

The most time demanding part.

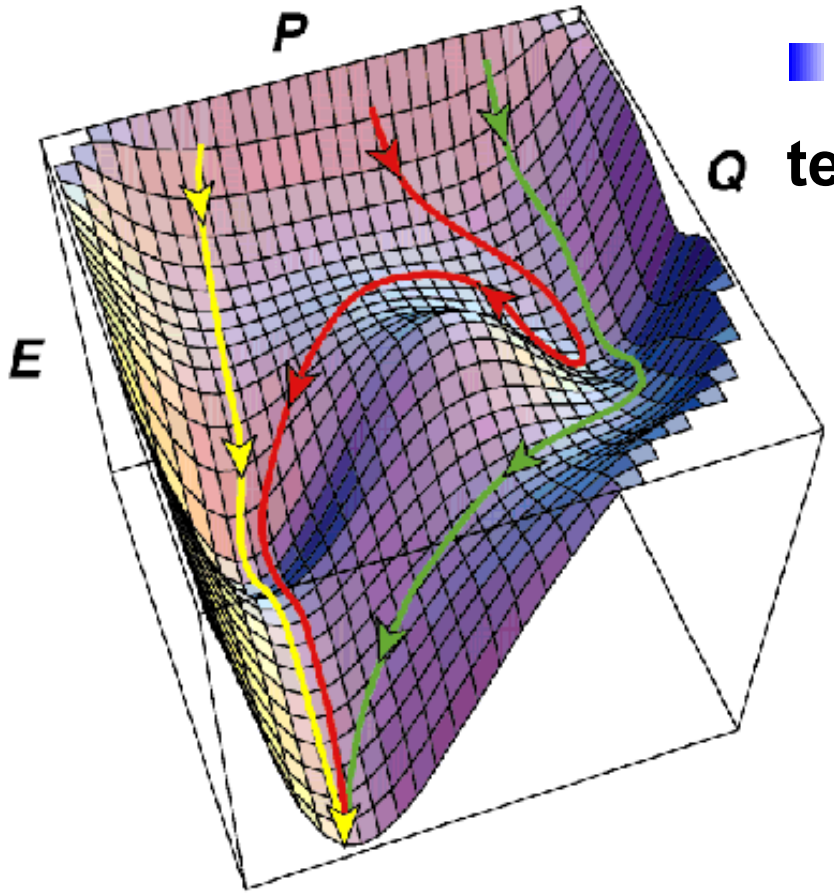
H-bonding term

Van der Waals term

Electrostatic term

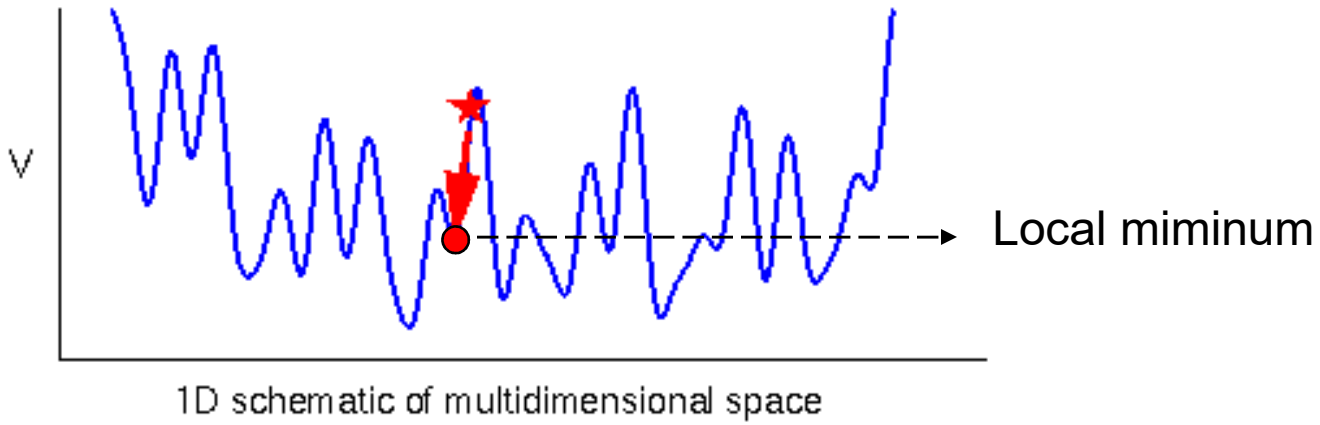
Search Potential Energy Surface

We are interested in minimum points on Potential Energy Surface (P



- **Conformational search techniques**
  - Energy Minimization
  - Monte Carlo
  - Molecular Dynamics
  - Others: Genetic Algorithm, Simulated Annealing

Energy Minimization



- Energy minimization
- Methods
  - First-order minimization: Steepest descent, Conjugate gradient minimization
  - Second derivative methods: Newton-Raphson method
  - Quasi-Newton methods: L-BFGS

Monte Carlo

- **Monte Carlo**

  In molecular simulations, 'Monte Carlo' is an importance sampling technique.

  1. Make random move and produce a new conformation
  2. Calculate the energy change ΔE for the new conformation
  3. Accept or reject the move based on the Metropolis criterion

$$P = \exp\left(-\frac{\Delta E}{kT}\right)$$

Boltzmann factor

If ΔE<0, P>1, accept new conformation;

Otherwise: P>rand(0,1), accept, else reject.

Monte Carlo

- **Monte Carlo (MC) algorithm**

  Generate initial structure R and calculate $E(R)$;

  Modify structure R to R' and calculate $E(R')$;

  Calculate $\Delta E = E(R') - E(R)$;

  IF $\Delta E<0$, then $R \leftarrow R'$;

  ELSE

  Generate random number RAND = rand(0,1);

  IF $\exp(-\Delta E/KT) >$ RAND, then $R \leftarrow R'$;

  ENDIF

  ENDIF

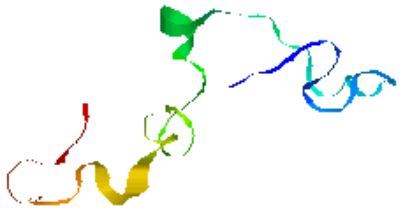
  Repeat for N steps;

- **Monte Carlo Minimization (MCM) algorithm**
- **Parallel Replica Exchange Monte Carlo algorithm**

Molecular Dynamics

- **Molecular Dynamics (MD)**

  MD simulates the Movements of all the particles in a molecular system by iteratively solving Newton's equations of motion.



MC view many frozen butterflies in a museum; MD watch the butterfly fly.

Molecular Dynamics

- **Algorithm**
  - For atom *i,* Newton's equation of motion is given by

$$F_i = m_i a_i \quad (1) \quad \Rightarrow \quad \mathbf{F}_i(t) = m_i \frac{\mathrm{d}^2 \mathbf{r}_i(t)}{\mathrm{d}t^2} \quad (2)$$

Here, $\mathbf{r}_i$ and m_i represent the position and mass of atom *i* and $\mathbf{F}_i(t)$ is the force on atom *i* at time *t*. $\mathbf{F}_i(t)$ can also be expressed as the gradient of the potential energy

$$\mathbf{F}_i = -\nabla_i V \quad (3) \quad \Rightarrow \quad -\nabla_i V = m_i \frac{\mathrm{d}^2 \mathbf{r}_i(t)}{\mathrm{d}t^2} \quad (4)$$

V is potential energy. Newton's equation of motion can then relate the derivative of the potential energy to the changes in position as a function of time.

Molecular Dynamics

- **Algorithm (continue)**
  - To obtain the movement trajectory of atom, numerous numerical algorithms have been developed for integrating the equations of motion. (Verlet algorithm, Leap-frog algorithm)

    Verlet algorithm

    The algorithm uses the positions and accelerations at time *t*, and the positions from the previous step $\mathbf{r}(t-\delta t)$ to calculate the new positions

    $$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \delta t^2 \mathbf{a}(t)$$

    Selection of time step

    Time step δt is approximately one order of magnitude smaller than the fastest motion

    Hydrogen vibration ~ 10 fs (10^{-15} s), time step = 1fs

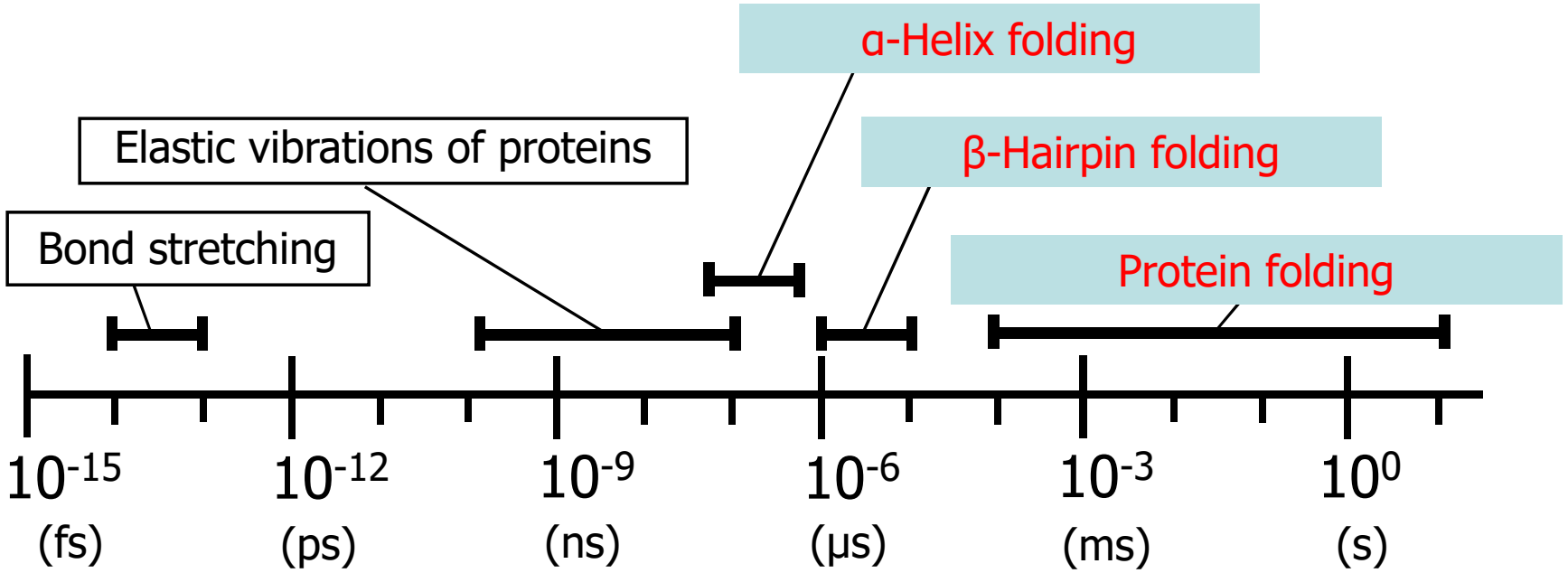
Molecular Dynamics

- **MD Software**
  - CHARMM (Chemistry at HARvard Molecular Mechanics) is a program for macromolecular simulations, including energy minimization, molecular dynamics and Monte Carlo simulations.
  - NAMD is a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems.

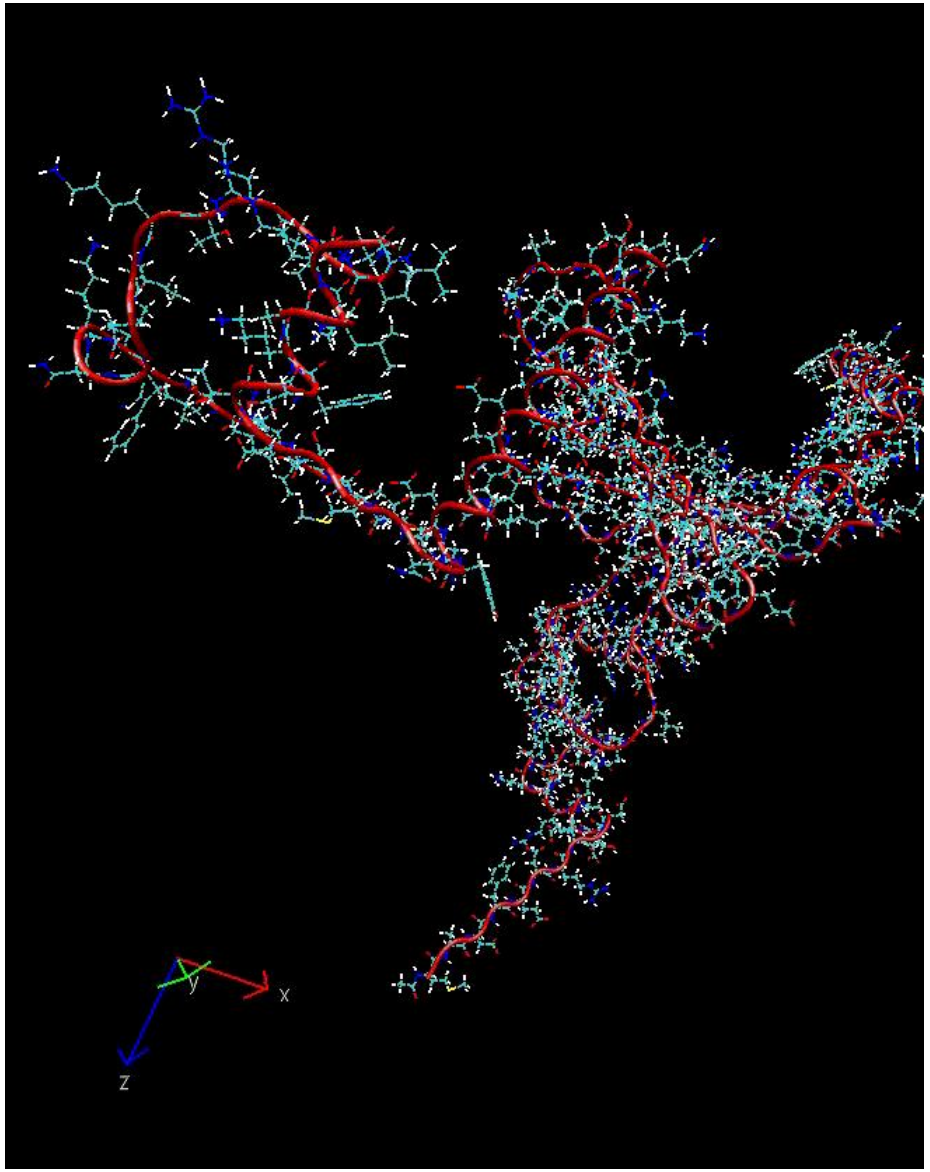
    http://www.ks.uiuc.edu/Research/namd/
- **Application in PSP**
  - Advantage: Deterministic; Provide details of the folding process
  - Limitation: The protein folding reactions take place at ms level, which is at the limit of accessible simulation times

It is still difficult to simulate a whole process of a protein folding using the conventional MD method.

Time Scales of Protein Motions and MD



MD Time Scale



MD is fun!

A small protein folding movie: simulated with NAMD/VMD

Other Conformational Search Algorithms

- **Global optimization algorithms**

  "Optimization" refers to trying to find the global energy minimum of a potential surface.

  - Genetic Algorithm (GA)
  - Simulated Annealing (SA)
  - Tabu Search (TS)
  - Ant Colony Optimization (ACO)

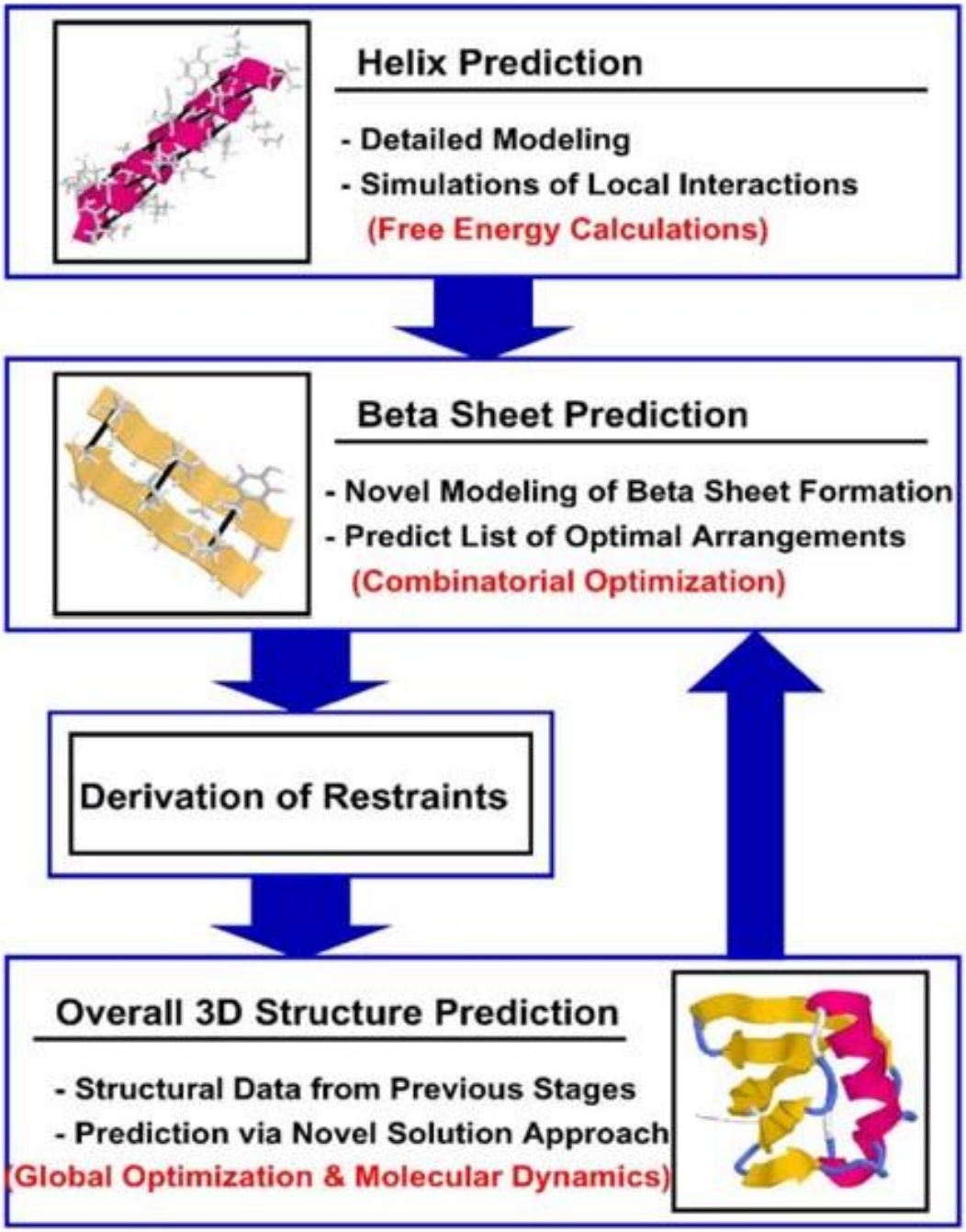
- A model system: Lennard Jones clusters

Applications of MM methods in PSP

- **Application in PSP**
  - Combination of several conformational search techniques
  - Recent developments
    - Simplified force field: united residue force field
    - Segment assembly

      Secondary structure prediction are quite reliable, so conformation can be produced by assemble the segments
- ***Ab initio* PSP software**
  - **Rosetta** is a five-stage fragment insertion Metropolis Monte Carlo method
  - **ASTRO-FOLD** is a combination of the deterministic αBB global optimization algorithm, and a Molecular Dynamics approach in torsion angle space
  - **LINUS** uses a Metropolis Monte Carlo algorithm and a simplified physics-based force field

ASTRO-FOLD


Helix Prediction
- Detailed Modeling
- Simulations of Local Interactions
(Free Energy Calculations)
Beta Sheet Prediction
- Novel Modeling of Beta Sheet Formation
- Predict List of Optimal Arrangements
(Combinatorial Optimization)
Derivation of Restraints
Overall 3D Structure Prediction
- Structural Data from Previous Stages
- Prediction via Novel Solution Approach
(Global Optimization & Molecular Dynamics)

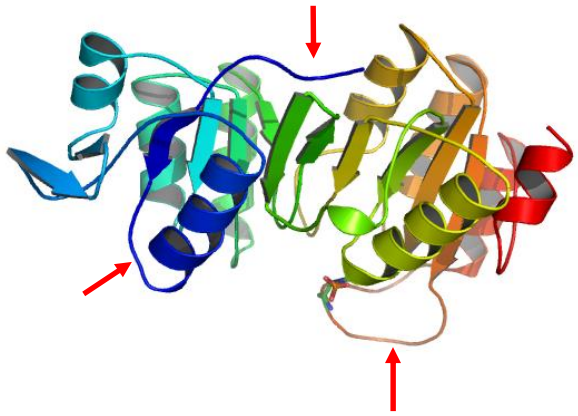
References

- Hardin C, *et. al.* Ab initio protein structure prediction. *Curr Opin Struct Biol.* 2002, 12(2):176-81.
- Floudas CA, *et. al.* Advances in protein structure prediction and de novo protein design: A review. *Chemical Engineering Science*, 2006, 61: 966-988.
- Klepeis JL, Floudas CA, ASTRO-FOLD: a combinatorial and global optimization framework for ab initio prediction of three dimensinal structures of proteins from the amino acid sequence, *Biophysical Journal*, 2003, 85: 2119-2146.

Ab Initio Protein Loop Prediction

- **Protein loop**

  Protein loops are polypeptides connecting more rigid structural elements of proteins like helices and strands.



- **Challenge in Loop Structure Prediction**
  - Loop is important to protein folding and protein function even though their size is small, usually <20 residues
  - Loops exhibit greater structural variability than helices and strands
  - Loop prediction is often a limiting factor on fold recognition methods

Ab Initio Protein Loop Prediction

- **Ab initio methods have recently received increased attention in the prediction of protein loop**

- **Potential energy function**

  Molecular mechanics force field is usually better than statistical potential in protein loop modeling.

- **Recent progress**
  - Dihedral angle sampling
  - Clustering
  - Select representative structures from ensembles

Ab Initio Loop Prediction Methods

- **Loopy**
  - Random tweak
  - Colony energy
- **Fiser's method**
  - MM methods:

    Physical energy function

    Energy Minimization + MD + SA
- **Forrest & woolf**
  - Predict membrane protein loop
  - MM methods: MC + MD

Review:

Floudas C.A. et al, Advances in protein structure prediction and de novo protein design: A review, *Chemical Engineering Science*, 2006, vol. 61, 966-988.

CLOOP: *Ab Initio* Loop Modeling Method

- CLOOP build all-atom ensemble of protein loop conformations (it is not a real protein loop prediction method)
- Paper

  Haiyan Jiang, Christian Blouin, *Ab Initio* Construction of All-atom Loop Conformations, *Journal of Molecular Modeling*, 2006, 12, 221-228.
- **CLOOP methods**
  - Energy function: CHARMM
  - Dihedral sampling
  - Potential smoothing technique
  - The designed minimization (DM) strategy
  - Divided loop conformation construction

The Energy Function of CHARMM Forcefield

- **CHARMM**

$$E_{CHARMm} = E_{bonds} + E_{UB} + E_{angle} + E_{dihe} + E_{imp} + E_{vdw} + E_{elec}$$

$$E_{bonds} = \sum_{bonds} k_b (b - b_0)^2 \qquad E_{dihe} = \sum_{dihe} k_\chi (1 + (\cos(n\chi - \delta))$$

$$E_{UB} = \sum_{UB} k_{UB} (S - S_0)^2 \qquad E_{imp} = \sum_{imp} k_{imp} (\varphi - \varphi_0)^2$$

$$E_{angle} = \sum_{angle} k_\theta (\theta - \theta_0)^2$$

$$E_{vdw} = \sum_{nonbond} \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right] \qquad E_{elec} = \sum_{nonbond} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

CLOOP

- **Dihedral sampling**
  - Loop main-chain dihedral ϕ and ψ are generated by sampling main-chain dihedral angles from a restrained ϕ/ψ set

    The restrained dihedral range has 11 pair of ϕ/ψ dihedral sub-ranges. It was obtained by adding 100 degree variation on each state of the 11 ϕ/ψ set developed by Mault and James for loop modeling.
  - Side chain conformations are built randomly.

CLOOP

- **Potential smoothing technique**

  A soft core potential provided in CHARMM software package was applied to smooth non-bonded interactions

$$E_{nonbond} = E_{nonbond}^{CHARMM} \qquad r \geq r_{soft}$$

$$E_{nonbonded} = k(r - r_{soft}) + E_{nonbonded}^{CHARMM} \qquad r < r_{soft}$$

r is the distance of the two interacting atoms

r_{soft} is the switching distance for the soft core potential

CLOOP

- **The designed minimization (DM) strategy**
  - Minimization methods:

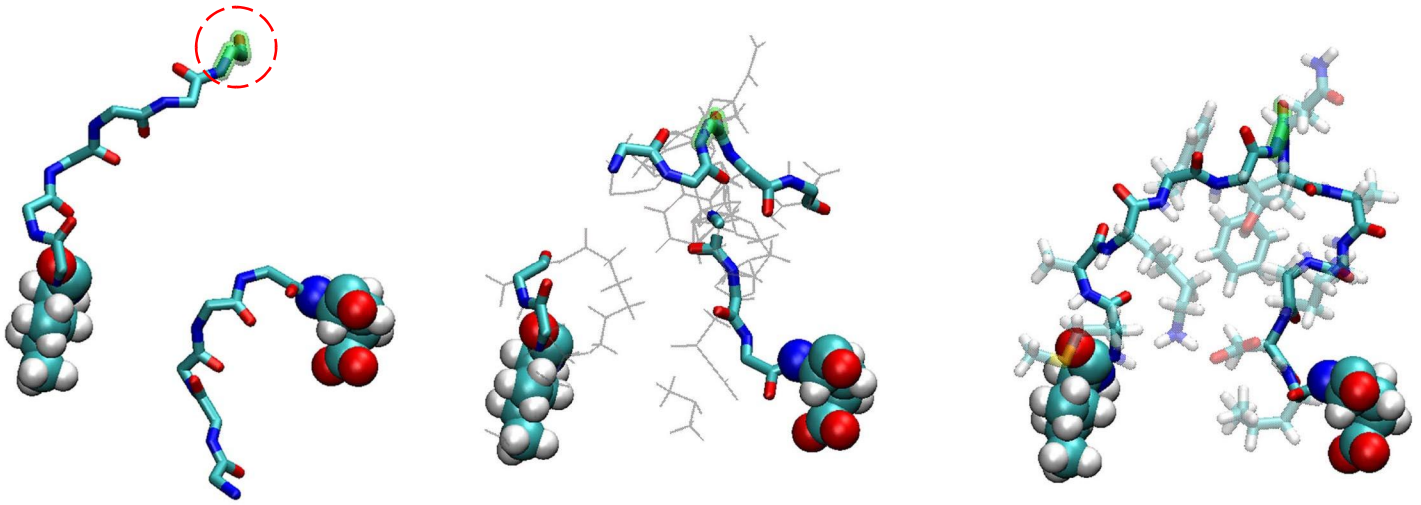
    steepest descent, conjugate gradient, and adopted basis Newton-Raphson minimization method
  - Two stages:

    1. Minimize the internal energy terms of loop conformations including bond, angle, dihedral, and improper

    2. The candidates were further minimized with the full CHARMM energy function including the van der Waal and electrostatic energy terms.

CLOOP

- **Divided loop conformation construction**



  - Generate position of middle residue
  - Build initial conformation of main chain with dihedral sampling
  - Build side chain conformation
  - Run DM and produce closed loop conformation

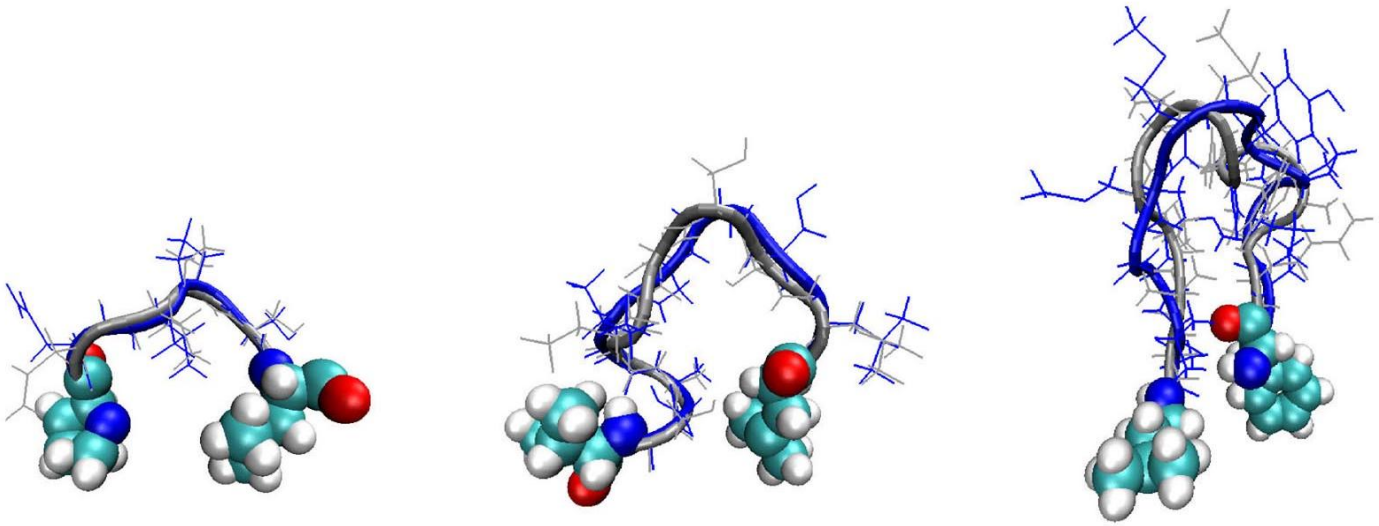
CLOOP

- **Performance of CLOOP**

CLOOP was applied to construct the conformations of 4, 8, and 12 residue long loops in Fiser's loop test set. The average main-chain root mean square deviations (RMSD) obtained in 1000 trials for the 10 different loops of each size are 0.33, 1.27 and 2.77 Å, respectively.

The performance of CLOOP was investigated in two ways. One is to calculate loop energy with a buffer region, and the other is loop only. The buffer region included a region extending up to 10 Å around the loop atoms. In energy minimization, only the loop atoms were allowed to move and all non-loop atoms include those in the buffer region were fixed.

Loop Conformations built by CLOOP



a. 1gpr_123-126 **b.** 135l_84-91 **c.** 1pmy_77-88

Performance of CLOOP

Table 4 Performance of CLOOP in 1,000 trials, the conformational energy of loop was calculated with a buffer region

Length 4		Length 8		Length 12	
Loop	$RMSD^{a}_{best}$	Loop	$RMSD_{best}$	Loop	$RMSD_{best}$
1aaj_82–85	0.28	135l_84–91	1.10	154l_153–164	3.13
1ads_99–102	0.47	1alc_34–41	1.62	1arp_201–212	2.64
1cbs_21–24	0.28	1btl_50–57	1.42	1ctm_9–20	3.00
1fkf_42–45	0.27	1cbs_55–62	1.24	1eco_35–46	2.60
1frd_59–62	0.42	1ddt_127–134	1.28	1ede_150–161	2.45
1gpr_123–126	0.22	1fnd_262–269	1.13	1ezm_122–133	2.42
1iab_100–103	0.31	1gky_72–79	1.17	1hfc_165–176	3.36
1mba_97–100	0.28	1iab_48–55	1.17	1msc_9–20	2.97
1nfp_37–40	0.38	1nar_192–199	1.32	1pbe_129–140	2.78
1pbe_117–120	0.35	1phf_85–92	1.23	1pmy_77–88	2.33
Average	0.33	Average	1.27	Average	2.77
Average $RMSD^{b}_{E}$	1.82	Average $RMSD_{E}$	4.04	Average $RMSD_{E}$	6.79
Average Time [c] (min)	23.6	Average Time (min)	49.0	Average Time (min)	70.1
Average N^{d}_{closed}	983	Average N_{closed}	922	Average N_{closed}	767

[a] Main-chain RMSD of the best conformation generated by CLOOP from X-ray structure
[b] Average main-chain RMSD of the conformational ensemble including all the good conformations
[c] Average computational time to generate 1,000 loop candidates for the test loop at certain length
[d] Average number of closed loop conformations

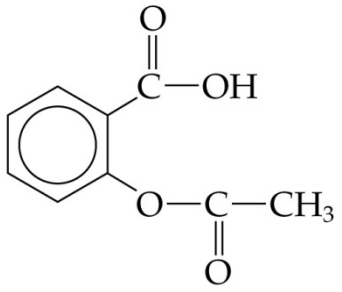
Conclusion

- CLOOP can be applied to build a good all-atom conformation ensemble of loops with size up to 12 residues.
- Good efficiency, CLOOP is faster than RAPPER
- The contribution of the protein to which a loop is attached (i.e. the ‘buffer region’) facilitates the discrimination of near-optimal loop structures.
- The soft core potentials and a DM strategy are effective techniques in building loop conformations.

UNIT-IV

Structure-based Drug Design

Pain relievers: aspirin



Acetylsalicylic acid
(Aspirin)

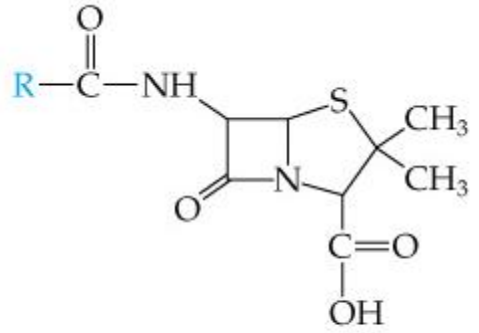
- Analgesic (pain reliever)
- Antipyritic (fever reducer)
- Anti-inflammatory
- Anticoagulent

History of Aspirin

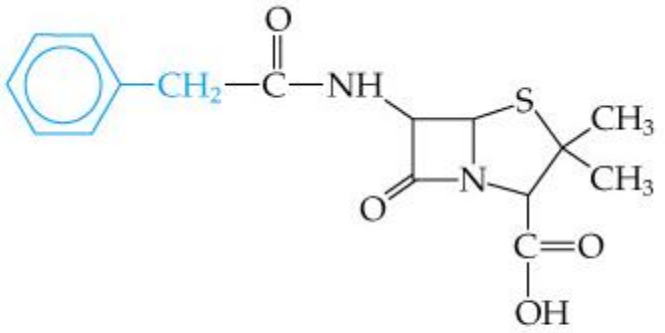
- Hippocratus: powder made from the bark and leaves of the willow tree to help heal headaches, pains and fevers
- Henri Leroux & Raffaele Piria: purification of active ingradient from the plant
- 1899 Hoffman: formulation and patent

Inhibits production of *prostaglandins* (pain messengers)

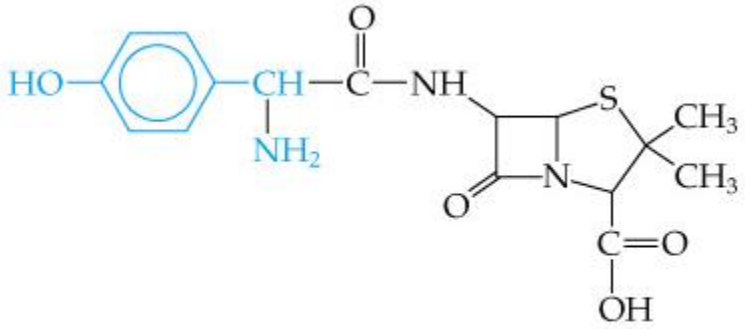
Antibacterial drugs: penicillins



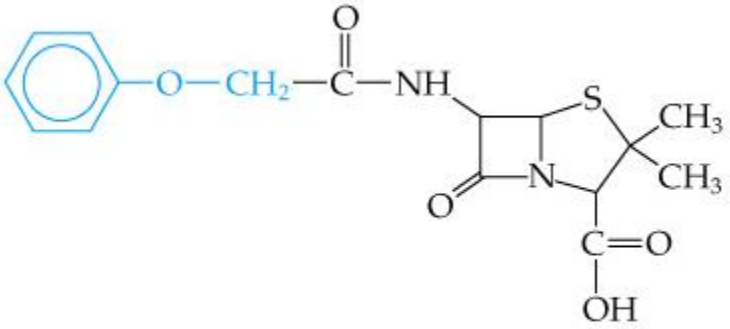
Penicillin
(a general formula)



Penicillin G



Amoxicillin



Penicillin V
(may be taken orally)

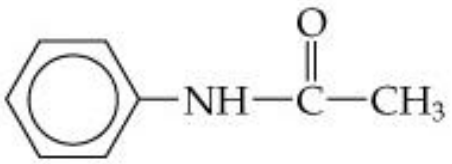
1941

Prevents crosslinking between proteins
and therefore cell wall synthesis (mucoproteins).

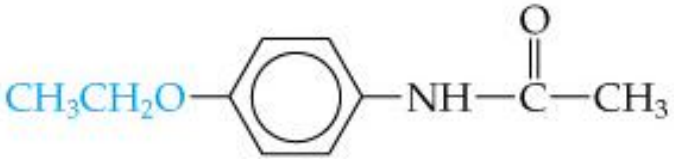
Aspirin substitutes

Now banned

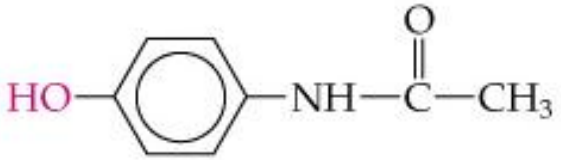
Tylenol



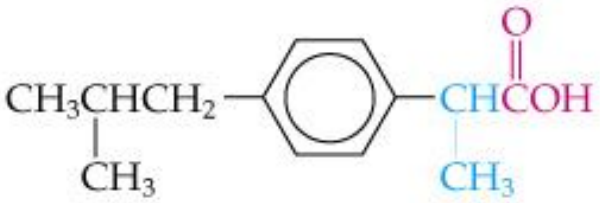
Acetanilide



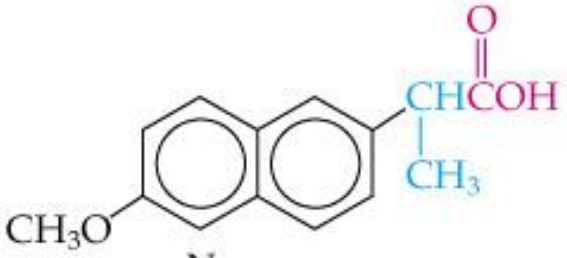
Phenacetin



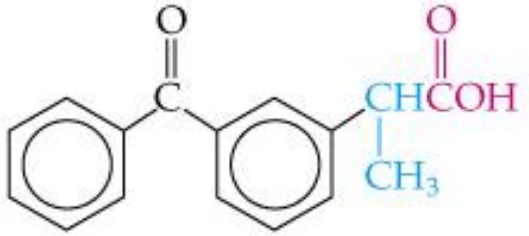
Acetaminophen



Ibuprofen



Naproxen



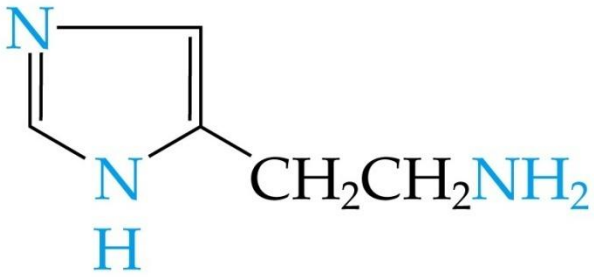
Ketoprofen

Advil

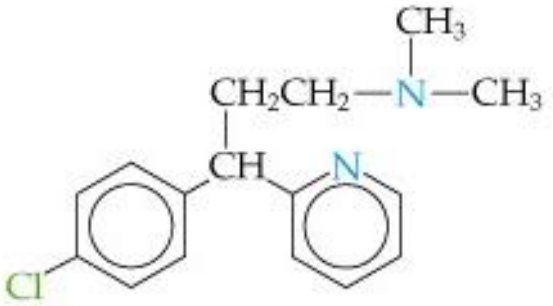
Aleve

Orudis KT

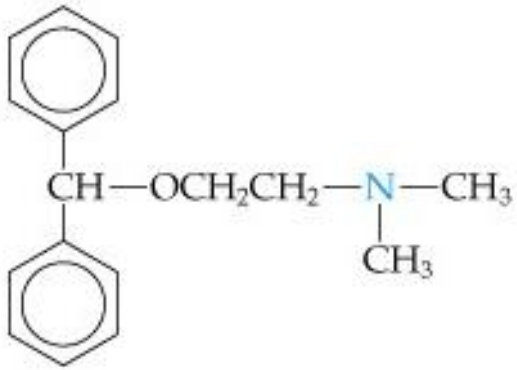
Antihistamines



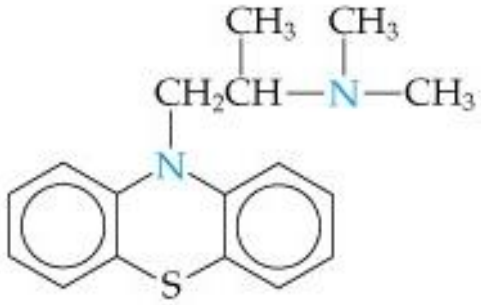
Histamine



Chlorpheniramine



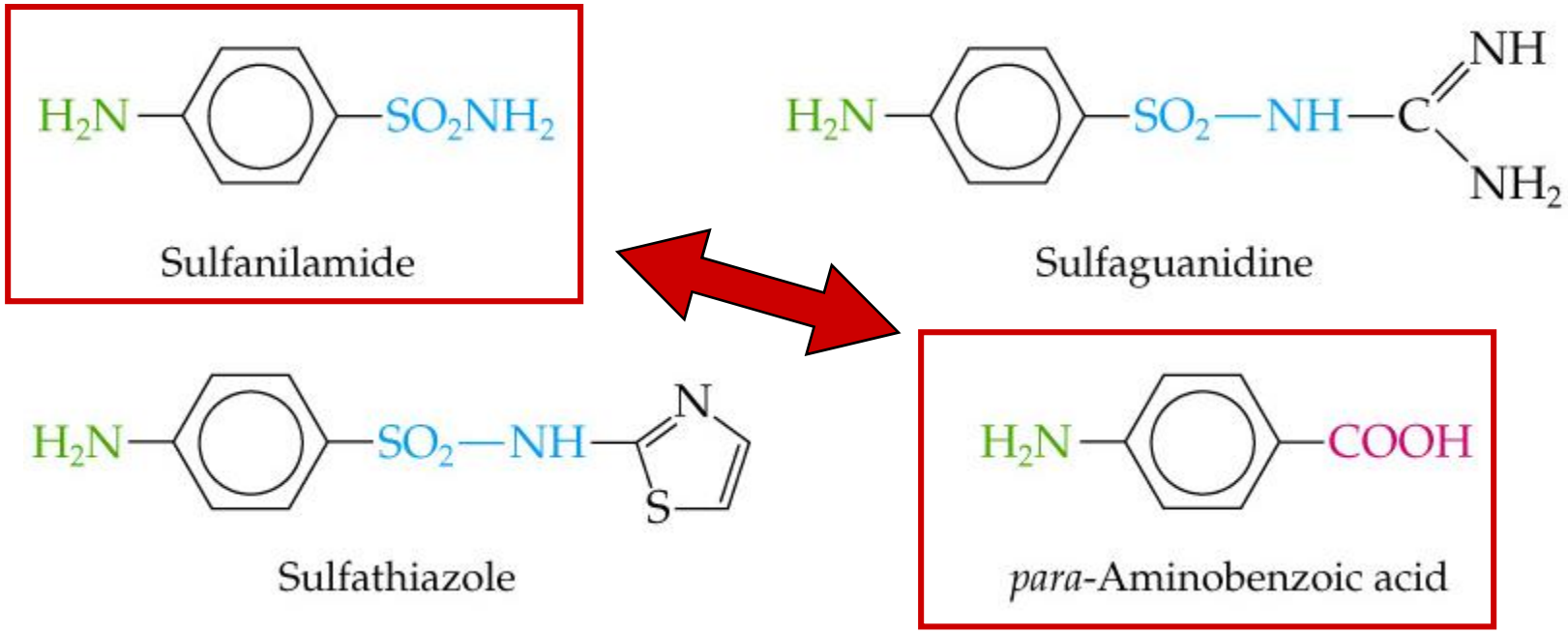
Diphenhydramine



Promethazine

(a) Antihistamines

Antibacterial drugs: sulfa drugs

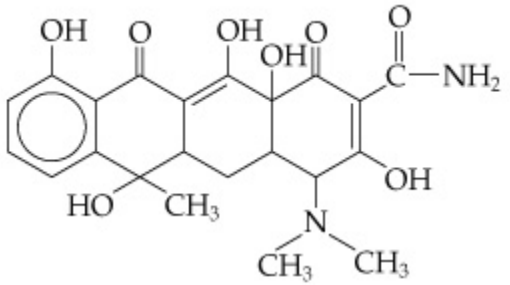


1935

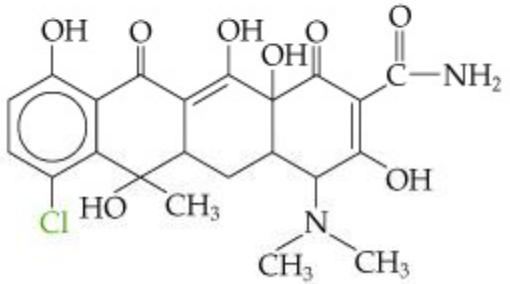
Chemical mimic-type poison for bacteria

Other antibacterial drugs

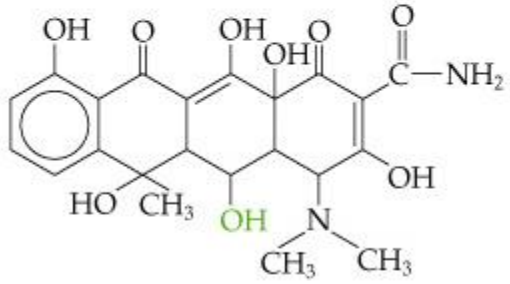
bind to bacterial ribosomes



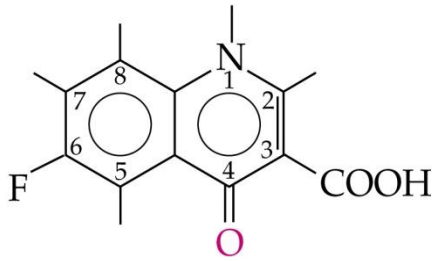
Tetracycline



Aureomycin
(chlorotetracycline)



Terramycin
(oxytetracycline)

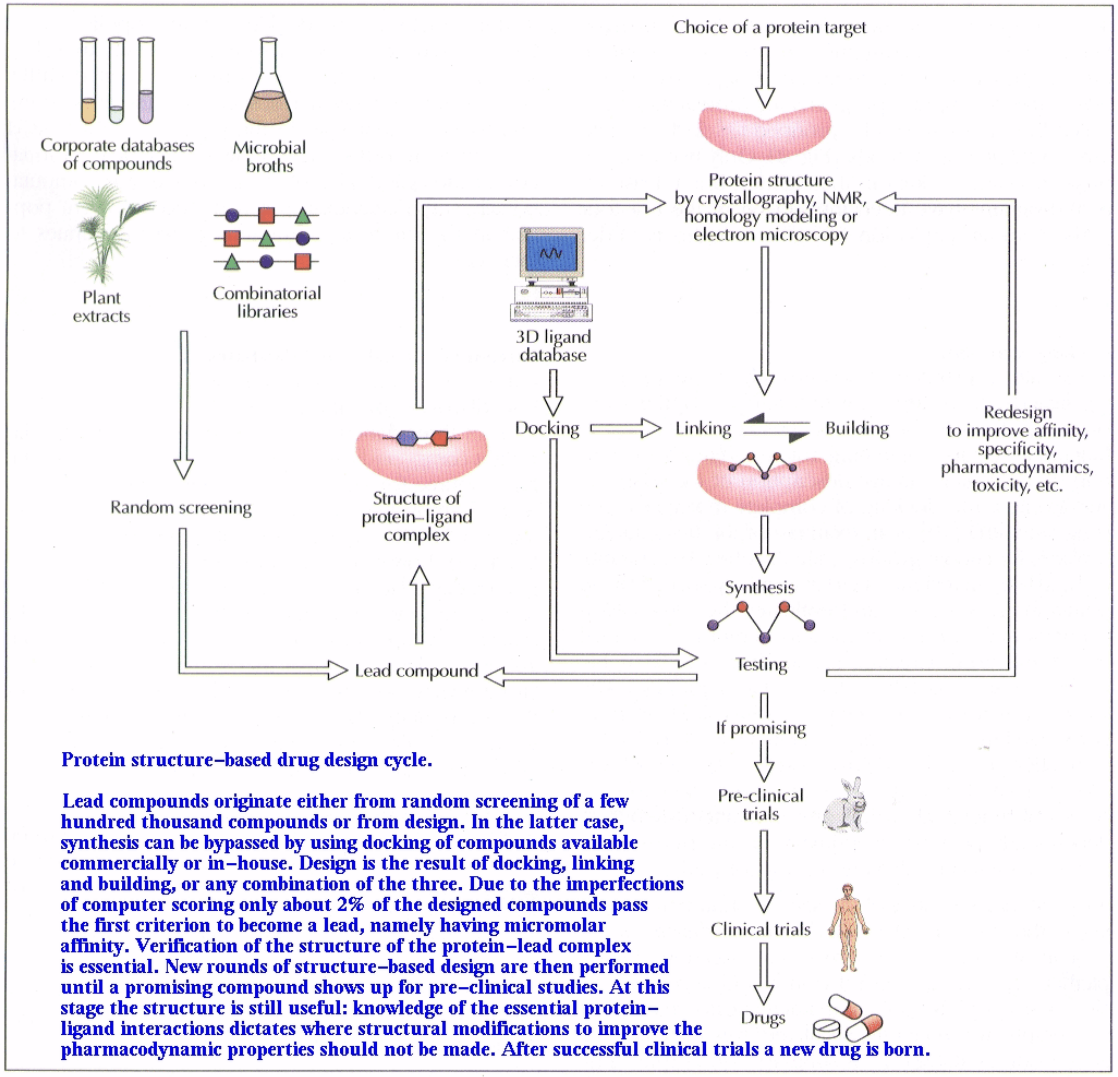


Fluoroquinolone

inhibits bacterial DNA replication

Structure-based Drug Design Cycle

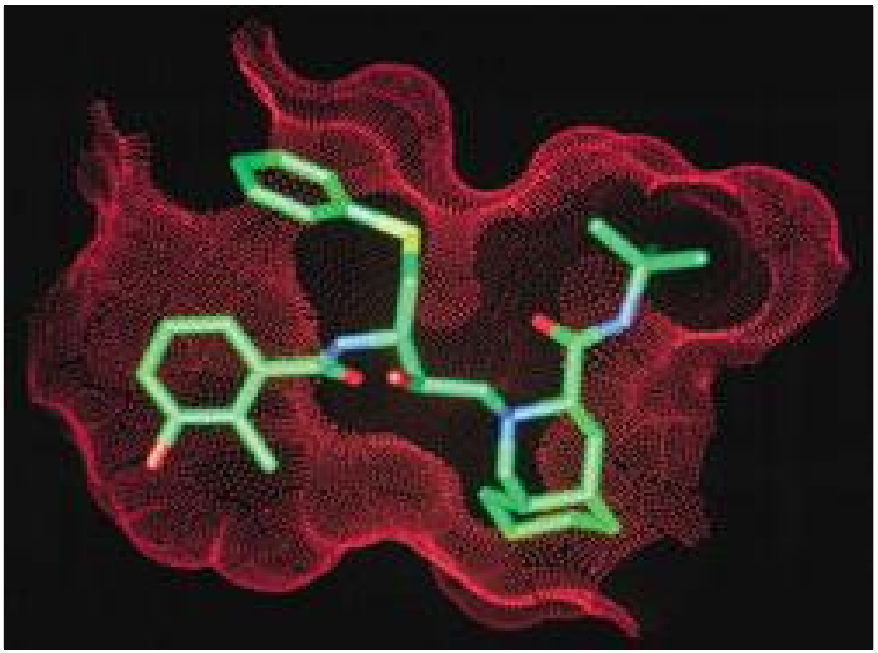
- Target identification and validation
- Assay development
- Virtual screening (VS)
- High throughput screening (HTS)
- Quantitative structure – activity relationship (QSAR) and refinement of compounds
- Characterization of prospective drugs
- Testing on animals for activity and side effects
- Clinical trials
- FDA approval



Protein structure–based drug design cycle.

Lead compounds originate either from random screening of a few hundred thousand compounds or from design. In the latter case, synthesis can be bypassed by using docking of compounds available commercially or in–house. Design is the result of docking, linking and building, or any combination of the three. Due to the imperfections of computer scoring only about 2% of the designed compounds pass the first criterion to become a lead, namely having micromolar affinity. Verification of the structure of the protein–lead complex is essential. New rounds of structure–based design are then performed until a promising compound shows up for pre–clinical studies. At this stage the structure is still useful: knowledge of the essential protein–ligand interactions dictates where structural modifications to improve the pharmacodynamic properties should not be made. After successful clinical trials a new drug is born.

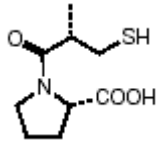
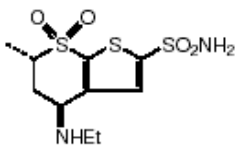
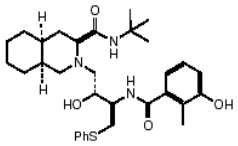
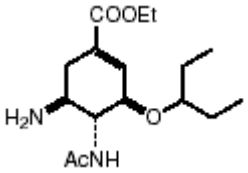
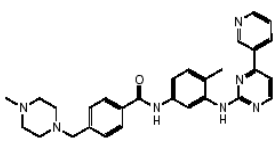
Drugs derived from structure-based approaches



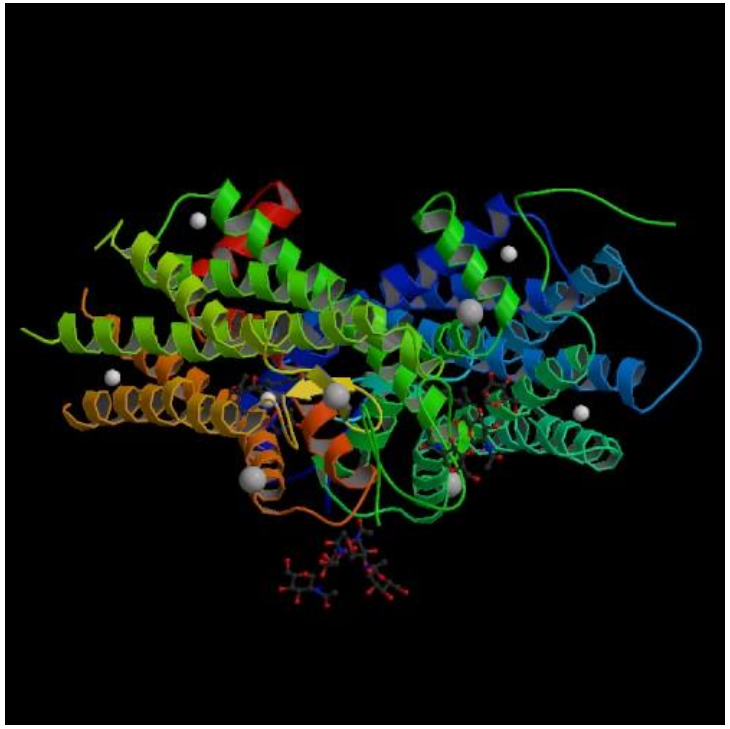
Nelfinavir in the active site of HIV-1 protease:

Agouron's AIDS drug nelfinavir (brand name Viracept) is one of the drugs on the market that can be traced directly to structure-based methods.

Drugs derived from structure-based approaches

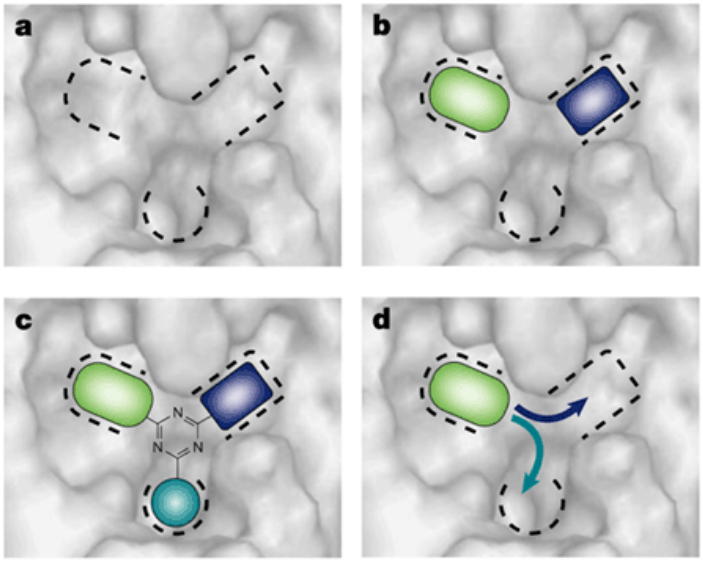
Capoten	Captopril		ACE	Hypertension	1981	Bristol-Myers Squibb
Trusopt	Dorzolamide		Carbonic anhydrase	Glaucoma	1995	Merck
Viracept	Nelfinavir		HIV protease	HIV/ AIDS	1999	Agouron (Pfizer) and Lilly
Tamiflu	Oseltamivir		Neuraminidase	Influenza	1999	Gilead and Roche
Gleevec	Imatinib		BCR- Abl	Chronic myelogenous leukaemia	2001	Novartis

Determination of Target Structure



Crystal structure of Rhodopsin: A G protein-coupled receptor.

Palczewski *et al. Science* (2000) **289**, 739- 45.



Nature Reviews | Drug Discovery

A. Binding site comprising three binding pockets

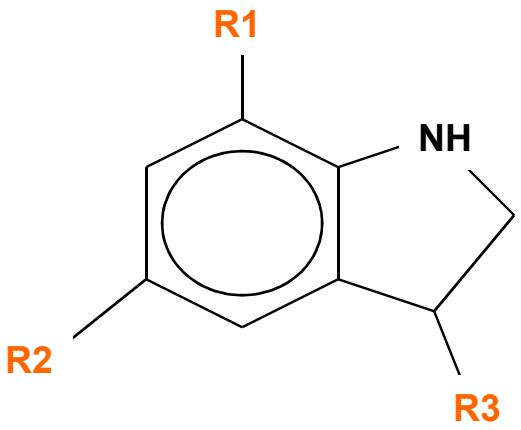
B. Crystallographic screening locates molecular fragments that bind to one, two or all three pockets

C. A lead compound is designed by organizing all three fragments around a core template

D. Growing out of a single fragment

Example Combinatorial Library

Scaffold



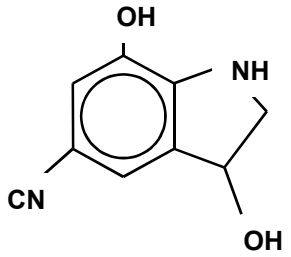
"R"-groups

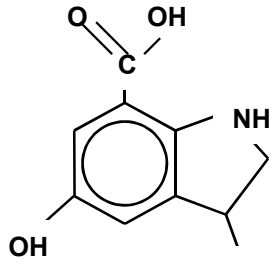
R1 = OH
OCH_3
NH_2
Cl
COOH

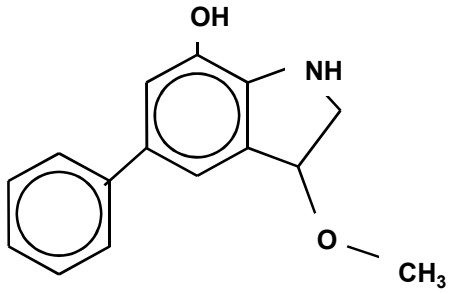
R2 = phenyl
OH
NH_2
Br
F
CN

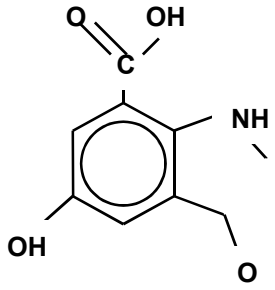
R3 = CF_3
NO_2
OCH_3
OH
phenoxy

Examples





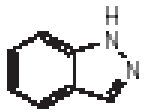
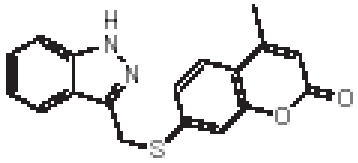
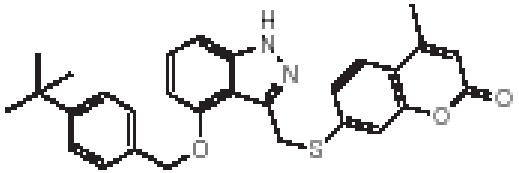
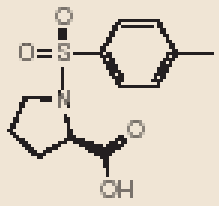
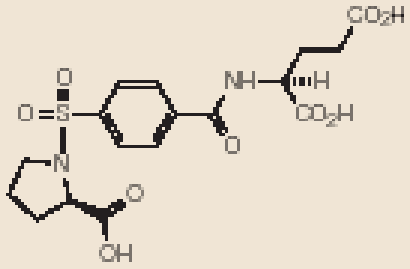
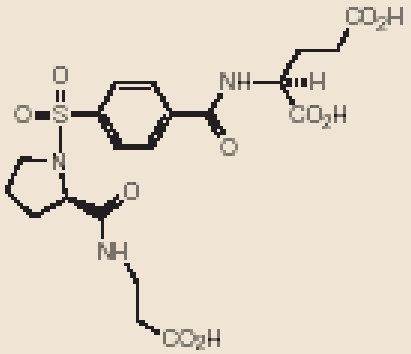
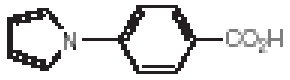
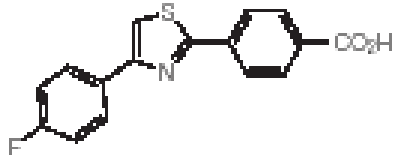
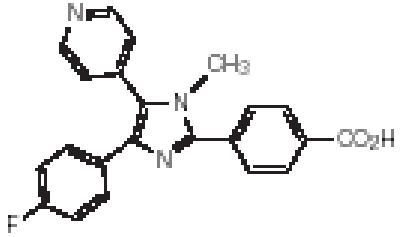




For this small library, the number of possible compounds is

5 x 6 x 5 = 150

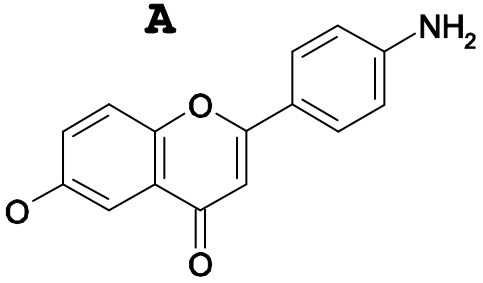
Lead Identification by Fragment Evolution

Entry	Target/method	Fragment	Evolved fragment	Lead
1	DNA gyrase[1]/ VS and SBD			
		K_i = 10 mM (by NMR) MNEC >250 μg per ml	MNEC = 8 μg per ml	MNEC = 30 ng per ml
2	Thymidylate synthase[19]/ tethering and SBD			
		IC_{50} = 1.1 mM	IC_{50} = 24 μM	IC_{50} = 330 nM
3	p38 kinase[2]/ NMR			
		K_d = 1 mM	K_d = 200 μM	K_i = 200 nM

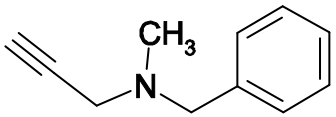
Similarity Paradox

Active

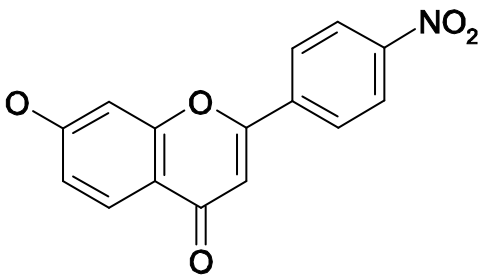
A



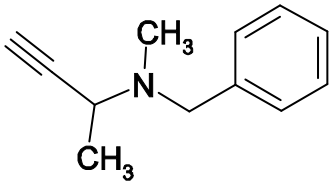
B



Inactive



C



D

Descriptors of Molecular Structure & Properties

- 1D-descriptors encode chemical composition & physicochemical properties
  - MW, $C_mO_nH_k$,hydrophobicity
- 2D-descriptors encode chemical topology
  - Connectivity indices, degree of branching, degree of flexibility, # of aromatic bonds
- 3D-descriptors encode 3D shape, volume, functionality, surface area
  - Pharmacophore – the spatial arrangement of chemical groups that determines its activity

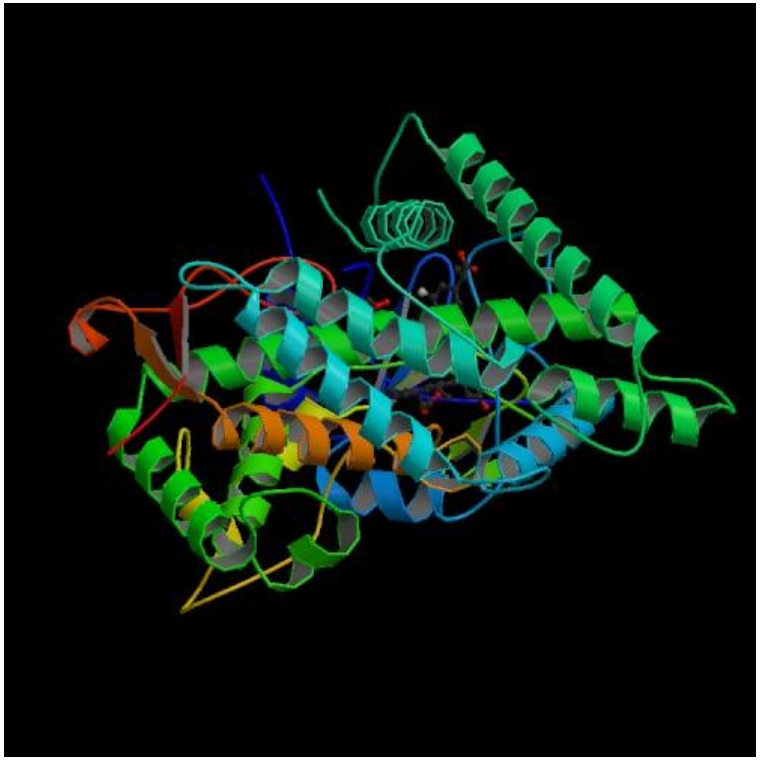
Lipinski Rule of Five (1997)

- *Poor absorption and permeation are more likely to occur when there are more than 5 hydrogen-bond donors, more than 10 hydrogen-bond acceptors, the molecular mass is greater than 500, or the log P value is greater than 5.*
- Further research studied a broader range of physicochemical and structural properties
- Related problems:
  - Compound toxicity
  - Compound mutagenicity
  - Blood-brain barrier penetration
  - Central nervous system activity

In Silico ADME Models

- Computational methods can predict compound properties important to ADME, e.g.
  - LogP, a liphophilicity measure
  - Solubility
  - Permeability
  - Cytochrome p450 metabolism
- Means estimates can be made for millions of compouds, helping reduce “attrition” – the failure rate of compounds in late stage

Can metabolism properties be modulated?



Structure of Cytochrome P450: responsible for primary metabolism of majority of drugs in human body -likely to herald a new era of structure-based design in the modulation of metabolic properties of drugs.

Ligand-based drug discovery

No a priori knowledge of the receptor

What information can we get from a few active compounds

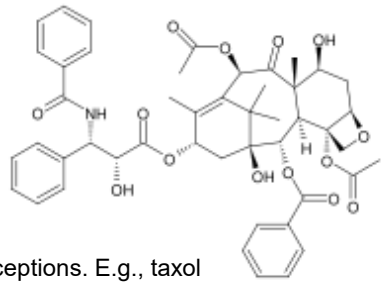
Drug-likeness: what makes a drug a drug?

Lipinski's "rule of 5": most drugs have less than two violation of the following rules:

Lipinski et al. Advanced Drug Delivery Reviews Volume 46, Issues 1-3, 1 March 2001, Pages 3-26

* ≤ 10 hydrogen bond acceptors
* ≤ 5 hydrogen bond donors
* Molecular weight ≤ 500

A partition coefficient log P ≤ 5



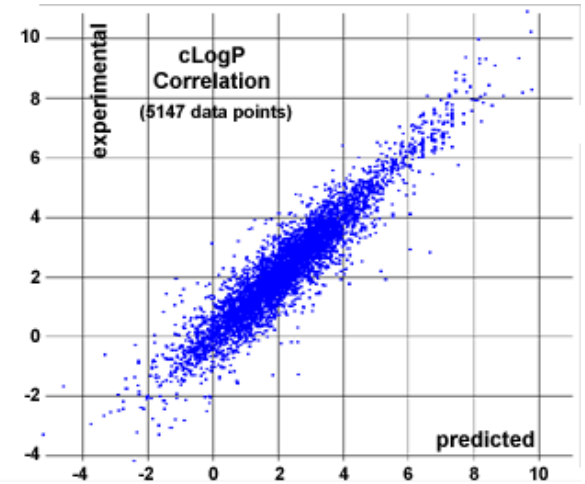
~30% exceptions. E.g., taxol

(non-original paper)" 2-8 rotatable bonds (entropy)

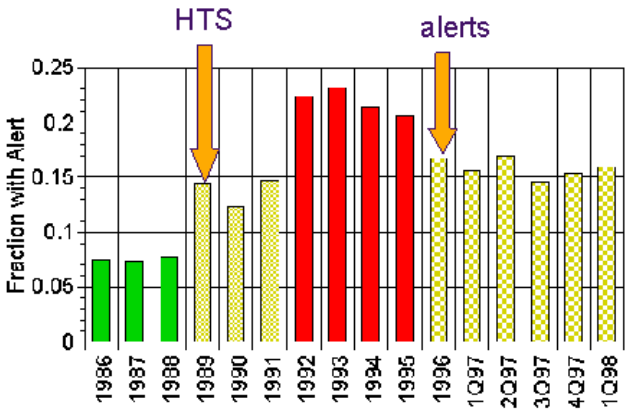
$$log\ P_{oct/wat} = log\left(\frac{[solute]_{octanol}}{[solute]^{un-ionized}_{water}}\right)$$

http://www.daylight.com/dayhtml/doc/clogp/#PCMsc1.2

benzene = 1.39
methanol = -.32
phenol = 1.4



The Age of Lipinski



http://cisrg.shef.ac.uk/shef2001/talks/davis/sld008.htm

- HTS lead generation biases chemistry

AstraZeneca

Hit development makes compounds larger and more complicated: violations of rule-of-5 criteria
Need to start from lower values that “drug-like”: development of lead-like criteria

Lead-like Profile

- Mwt 200-350
  - optimisation adds ca. 100
- logP 1-3
  - optimisation may increase by 1-2 logunits
- single charge
  - positive charge preferred
  - secondary or tertiary amine

1998: less than 600 solid compounds with mwt <250 and clogP <2
1999: 3000 added by purchase. Synthesis added >30000

AstraZeneca

Today:
Design lead-like libraries for vHTS

Even screen “probes”, scaffolds

QSAR:Build-up on known actives.

About 30 actives, a few log units in activity

Correlates physico-chemical properties of ligands with biological activity

Calculation of Molecular descriptors:

http://www.chemcomp.com/journal/descr.htm

a_aro	Number of aromatic atoms.
a_count	Number of atoms (including implicit hydrogens)
a_heavy	Number of heavy atoms
mW	Molecular weight

physical properties

apol	Sum of the atomic polarizabilities (including implicit hydrogens) with polarizabilities taken from [CRC 1994].
bpol	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens)
with	polarizabilities taken from [CRC 1994].
FCharge	Total charge of the molecule (sum of formal charges).
mr	Molecular refractivity (including implicit hydrogens). This property is calculated from an 11 descriptor linear model [MREF 1998] with r2 = 0.997, RMSE = 0.168 on 1,947 small molecules.

esoterism

The Subdivided Surface Areas are descriptors based on an approximate accessible van der Waals surface area calculation for each atom, vi along with some other atomic property, pi. The vi is calculated using a connection table approximation. Each descriptor in a series is defined to be the sum of the vi over all atoms, i such that pi is in a specified range (a,b].

In the descriptions to follow, Li denotes the contribution to logP(o/w) for atom i as calculated in the SlogP descriptor [Crippen 1999]. Ri denotes the contribution to Molar Refractivity for atom i as calculated in the SMR descriptor [Crippen 1999]. The ranges were determined by percentile subdivision over a large collection of compounds.

Code	Description
SlogP_VSA0	Sum of vi such that Li <= -0.4.
SlogP_VSA1	Sum of vi such that Li is in (-0.4,-0.2].
SlogP_VSA2	Sum of vi such that Li is in (-0.2,0].
SlogP_VSA3	Sum of vi such that Li is in (0,0.1].
SlogP_VSA4	Sum of vi such that Li is in (0.1,0.15].
SlogP_VSA5	Sum of vi such that Li is in (0.15,0.20].
SlogP_VSA6	Sum of vi such that Li is in (0.20,0.25].

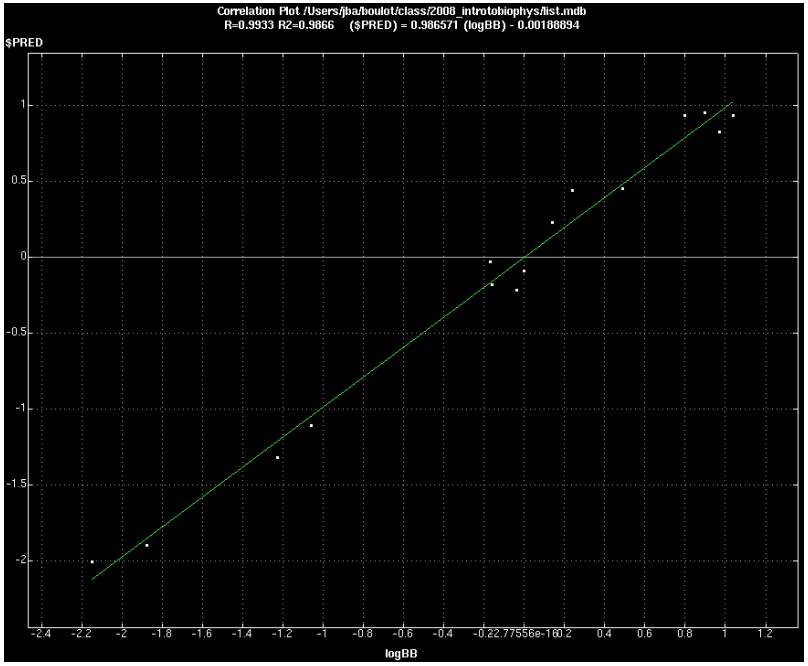
delirium tremens

balabanJ	Balaban's connectivity topological index [Balaban 1982]
weinerPol	Wiener polarity number: half the sum of all the distance matrix entries with a value of 3 as defined in [Balaban 1979].

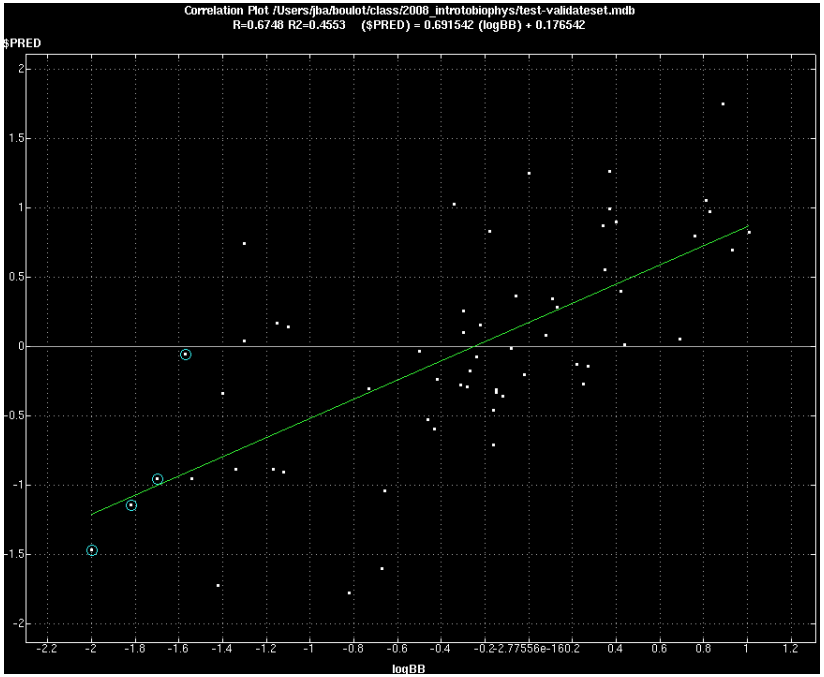
Drug optimisation : qsar

Exemple with 75 compounds.
Activity = log (blood/brain barrier)

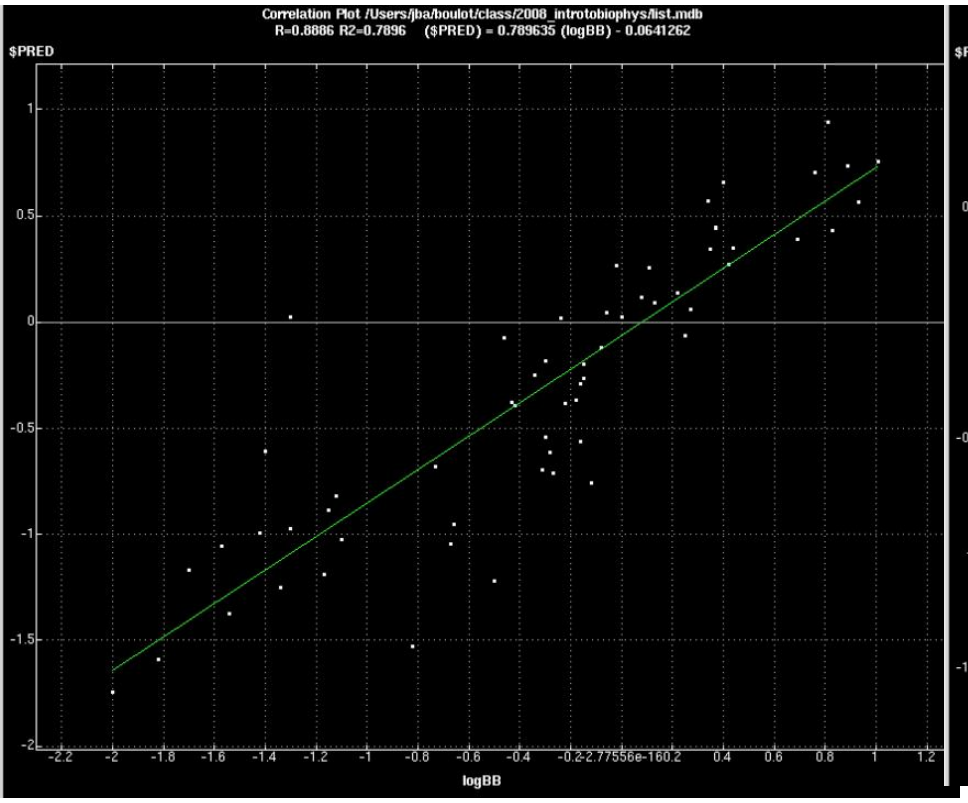
logBB =
0.07824
-0.00000 * petitjean
-0.00014 * diameter
-0.00000 * petitjeanSC
-0.00007 * radius
-0.00003 * VDistEq
-0.00003 * VDistMa
-0.00009 * weinerPath
-0.00023 * weinerPol
-0.00001 * BCUT_PEOE_0
+0.00004 * BCUT_PEOE_1
-0.00003 * BCUT_PEOE_2
+0.00002 * BCUT_PEOE_3
-0.00001 * BCUT_SLOGP_0
+0.00004 * BCUT_SLOGP_1
-0.00001 * BCUT_SLOGP_2
+0.00003 * BCUT_SLOGP_3
-0.00002 * BCUT_SMR_0
+0.00002 * BCUT_SMR_1
-0.00003 * BCUT_SMR_2
+0.00001 * BCUT_SMR_3

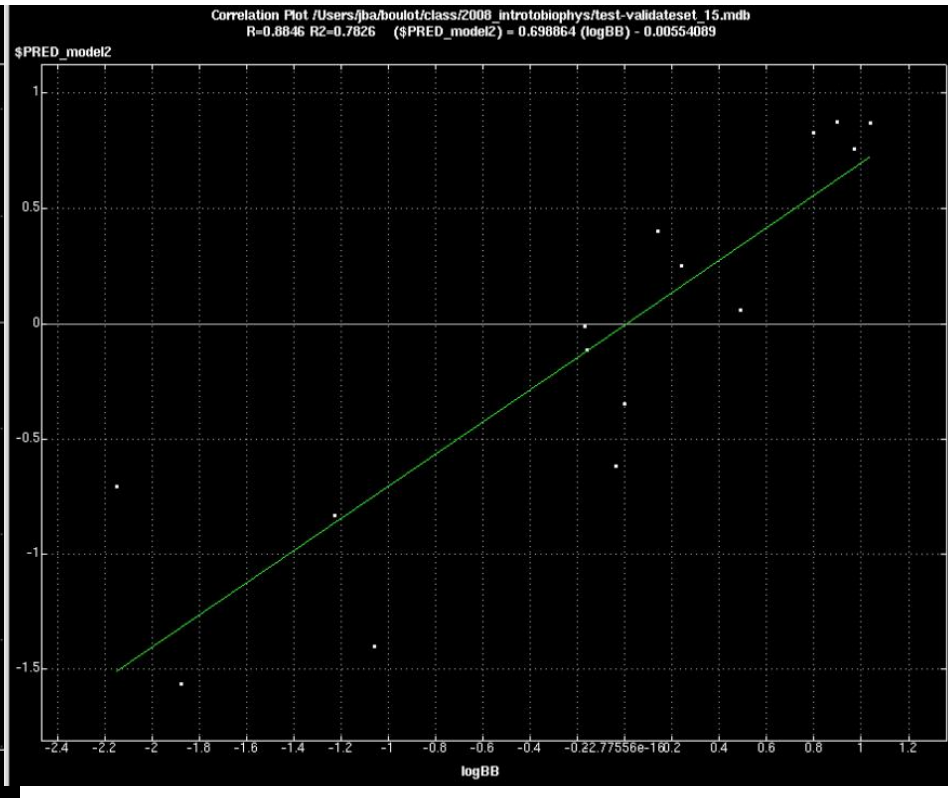


15 learning set



60 applications

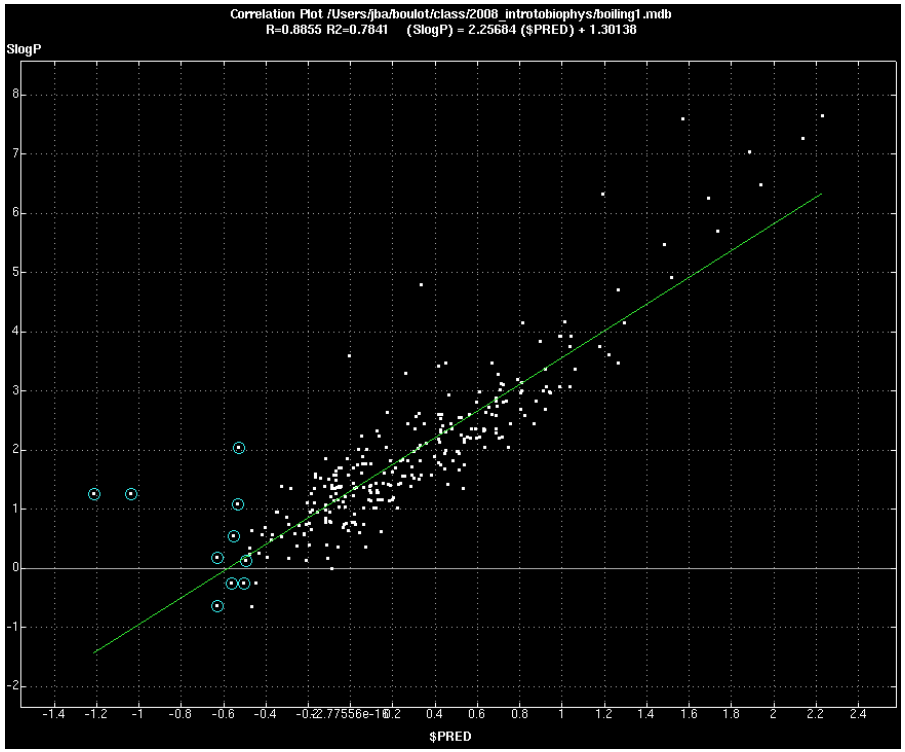


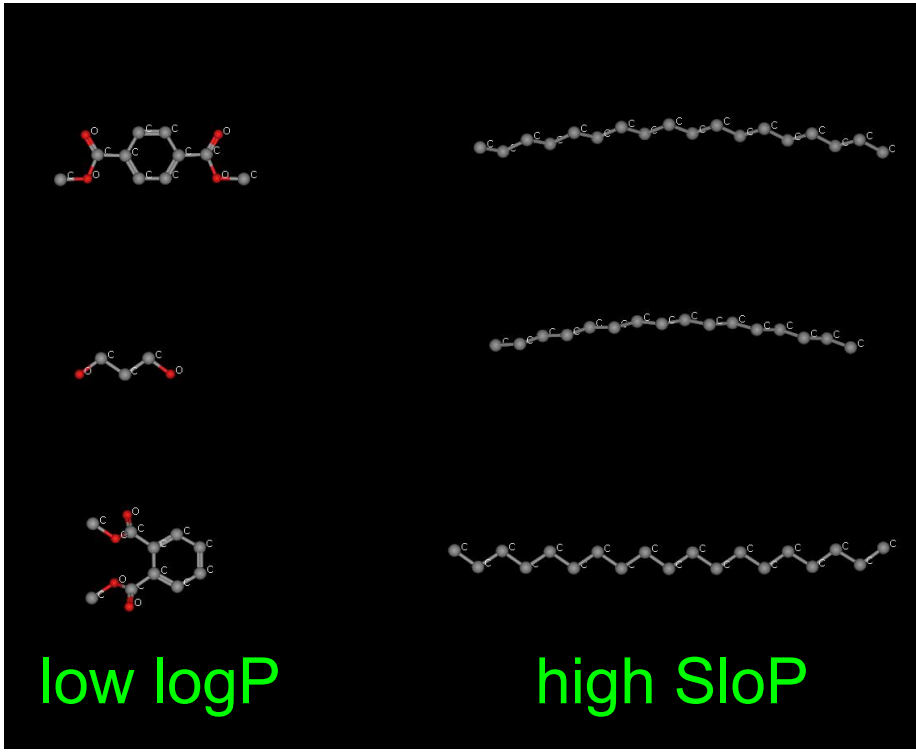


60 learning set

15 applications

Apply model on larger database

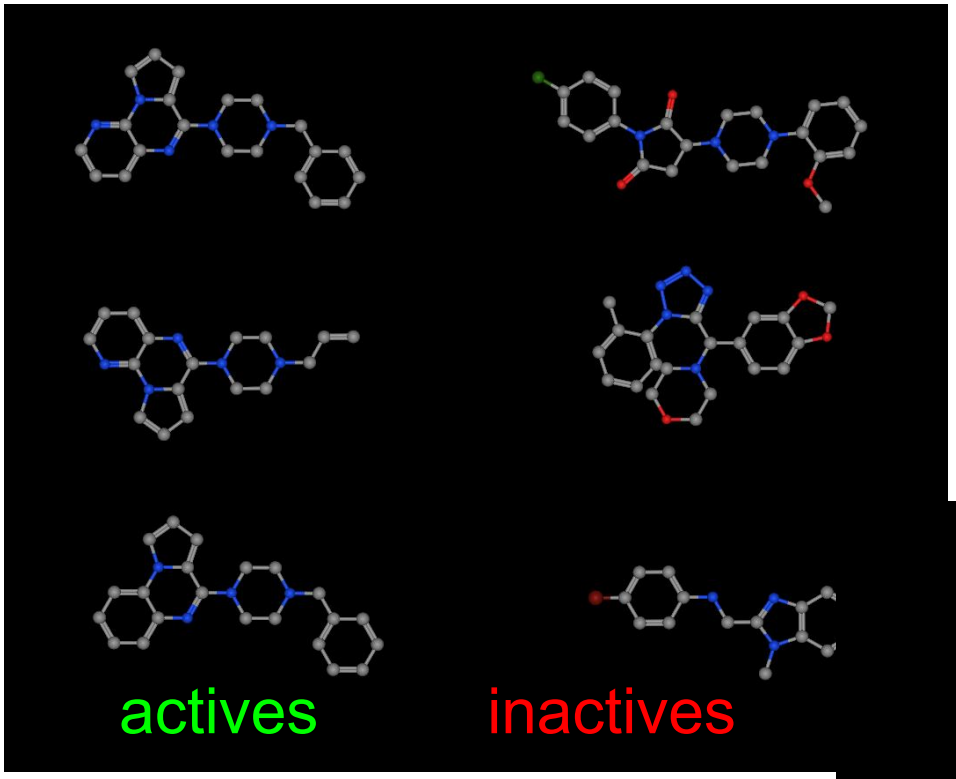


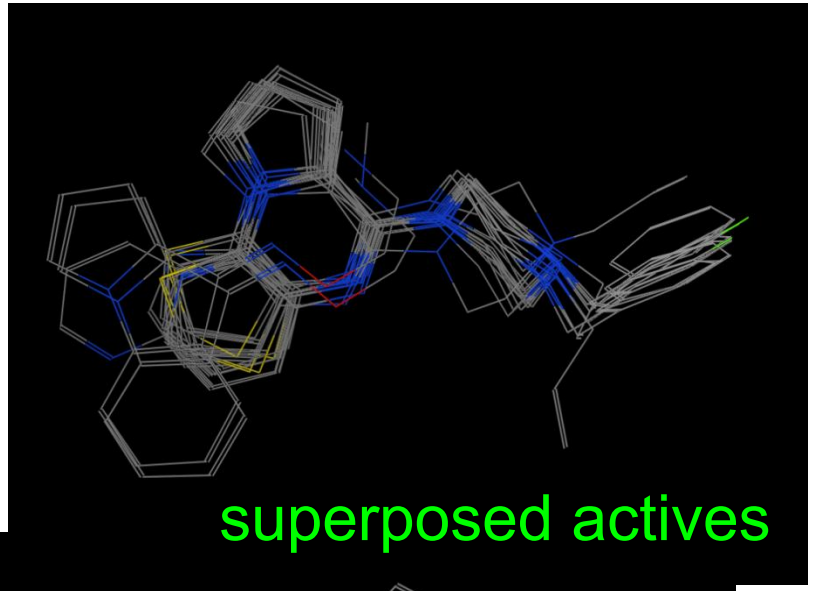


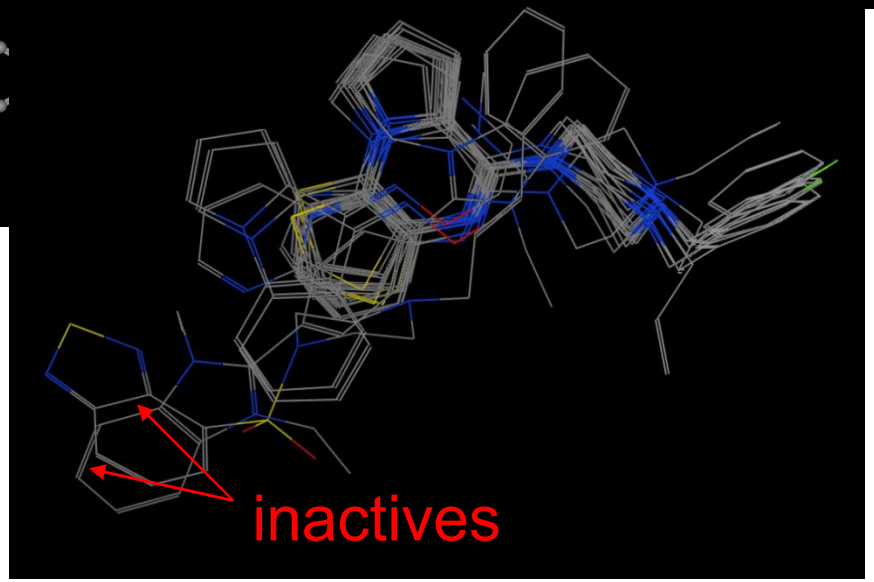
Would screen down the # of molecules to consider further for , e.g, synthesis, purchase, biology screening, structure-based screening etc…

Non descriptors-based : pharmacophore approaches

Ex: serotonine inhibitors

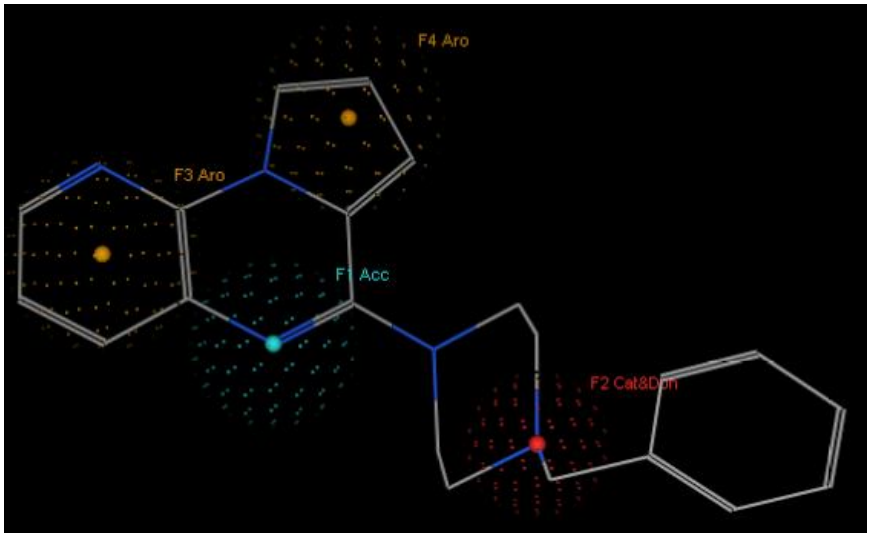


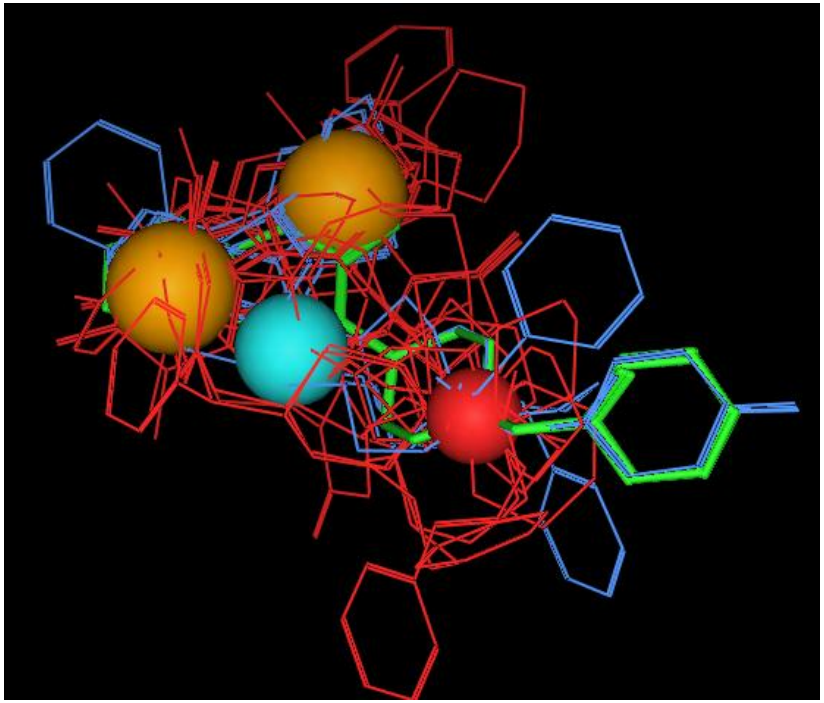




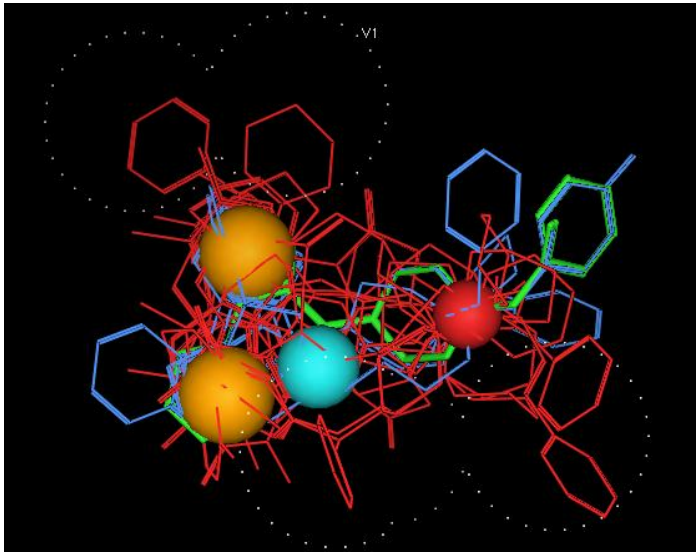
Active superimpose "better" Can we quantify & rationalize that?

From aligned molecules : define pharmacophores





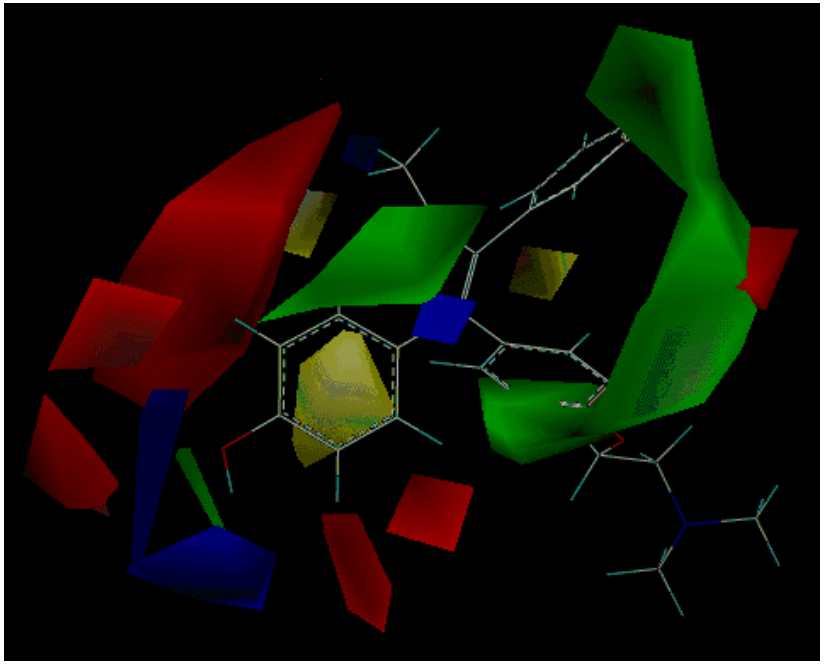
Consensus among most actives



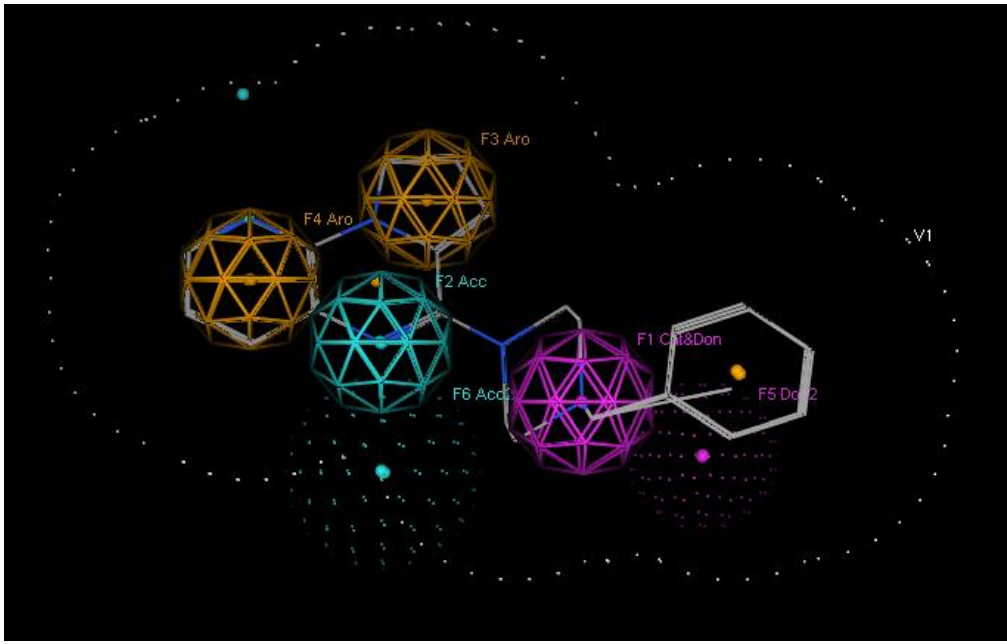
Testing and further development of pharmacophore model

QuickTime™ and a
decompressor
are needed to see this picture.

COMFA (“inverse” pharmacophore



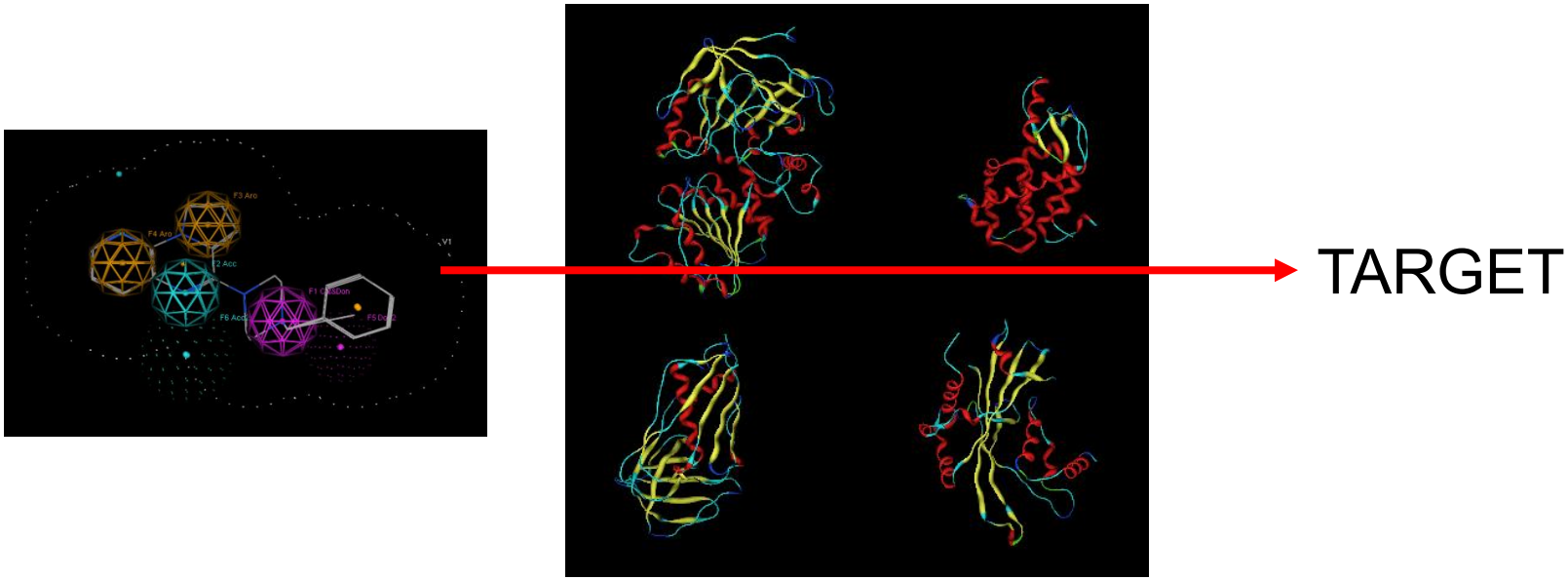
Pharmacophore of receptor



Green: Steric bulk favored
Yellow: Steric bulk disfavored
Blue: Positive charge and H-bond donors favored
Negative charge and H-bond acceptors disfavored
Red: Negative charge and H-bond acceptors favored
Positive charge and H-bond donors disfavore

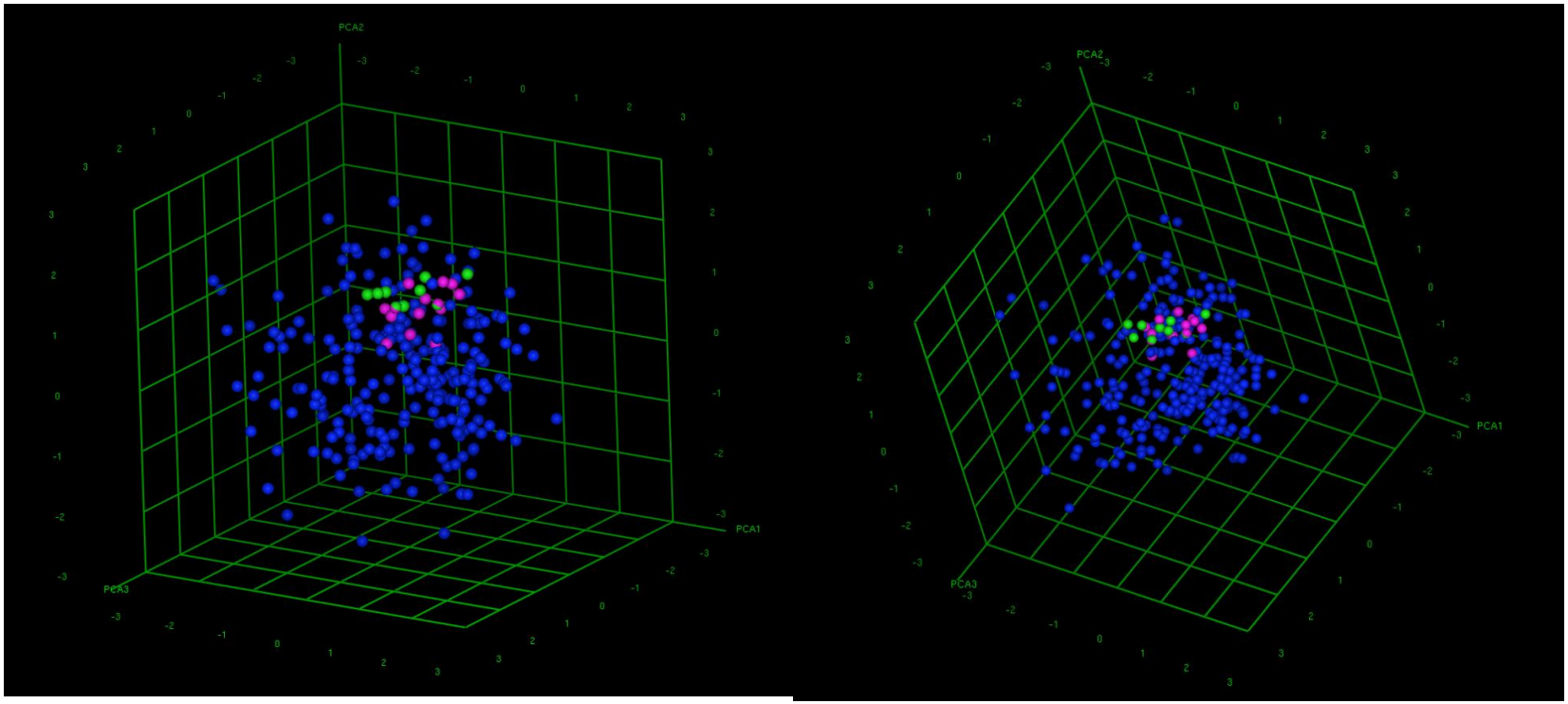
Use of receptor pharmacophore for receptor-selection: what protein could accommodate a given ligand.

Chemical Biology , target discovery , proteome exploration etc.…



Chemical space and molecular similarity/diversity

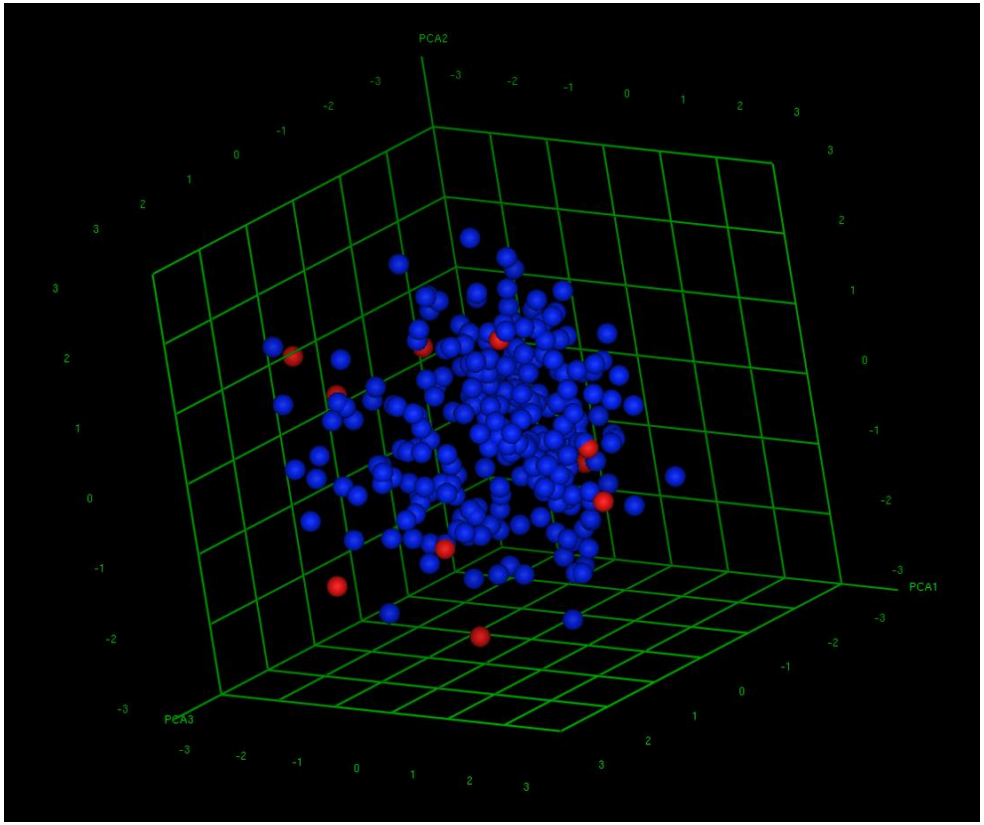
275 molecules, about 25 actives



3D representation of 5D PCA reduction of receptor-based QSAR

green: most active, purple: moderate, blue: inactive

diversity



In a large database, select compounds that exhibit most differences in properties: maximize chemical diversity

Diversity -> focused. Chemical-space-based drug discovery:
O Fishing with a large net in a large lake, use a smaller net where fish is found

UNIT-V

Receptorology

Drug receptor interaction

RECEPTORS

A RECEPTOR IS THE SPECIFIC CHEMICAL CONSTITUENT OF THE CELL WITH WHICH A DRUG INTERACTS TO PRODUCE IT'S PHARMACOLOGICAL EFFECTS.

Most of the receptors are Protein, but in some cases nucleic acid also acts as a receptor

Receptor : **Any cellular macromolecule that a drug binds to initiate its effects.**

Drug : **A chemical substance that interacts with a biological system to produce a physiologic effect. All drugs are chemicals but not all chemicals are drugs.**

The ability to bind to a receptor is mediated by the chemical structure of the drug that allows it to Interact with complementary surfaces on the receptor.

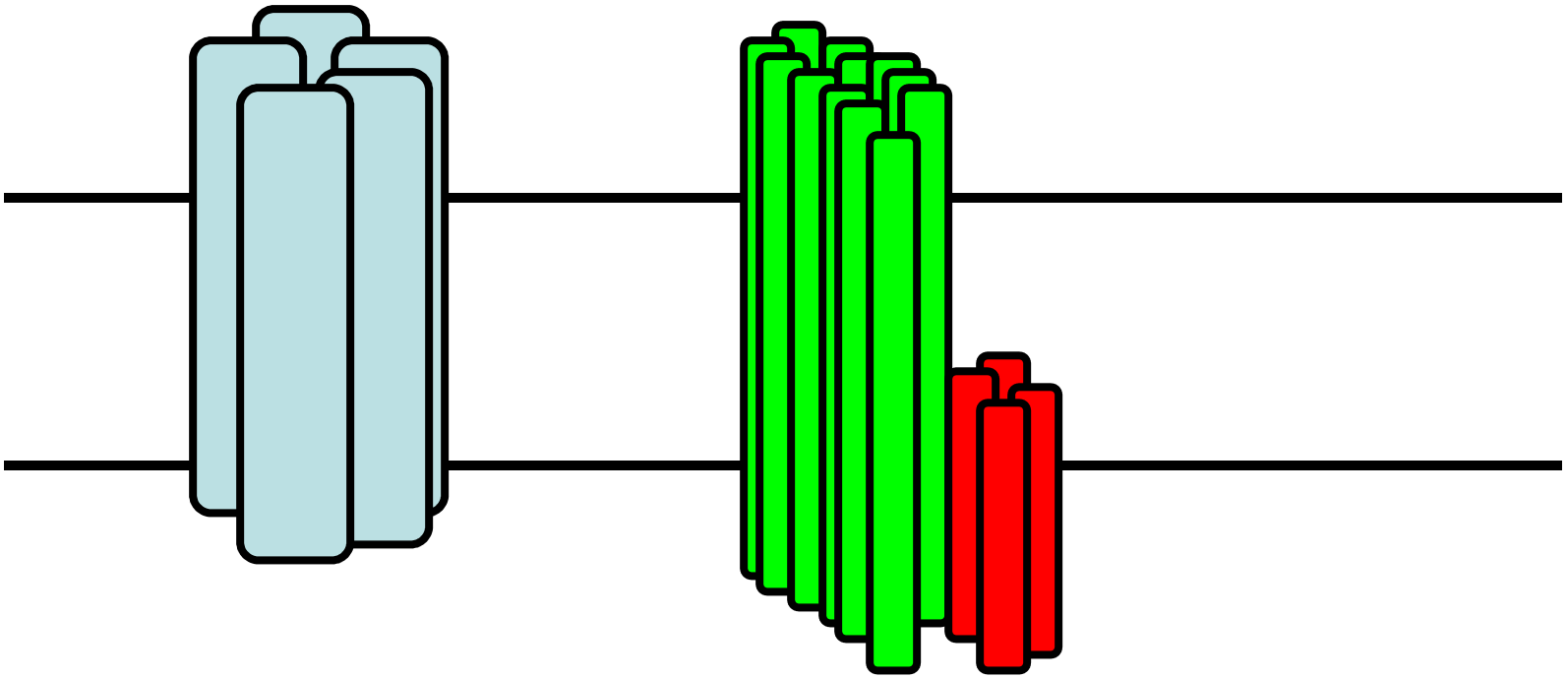
Types of interaction

- Hydrogen bond
- Vander Waals
- Electrostatic interaction
- Hydrophobic interaction
- Dipole-dipole interaction
- Ion – dipole interaction

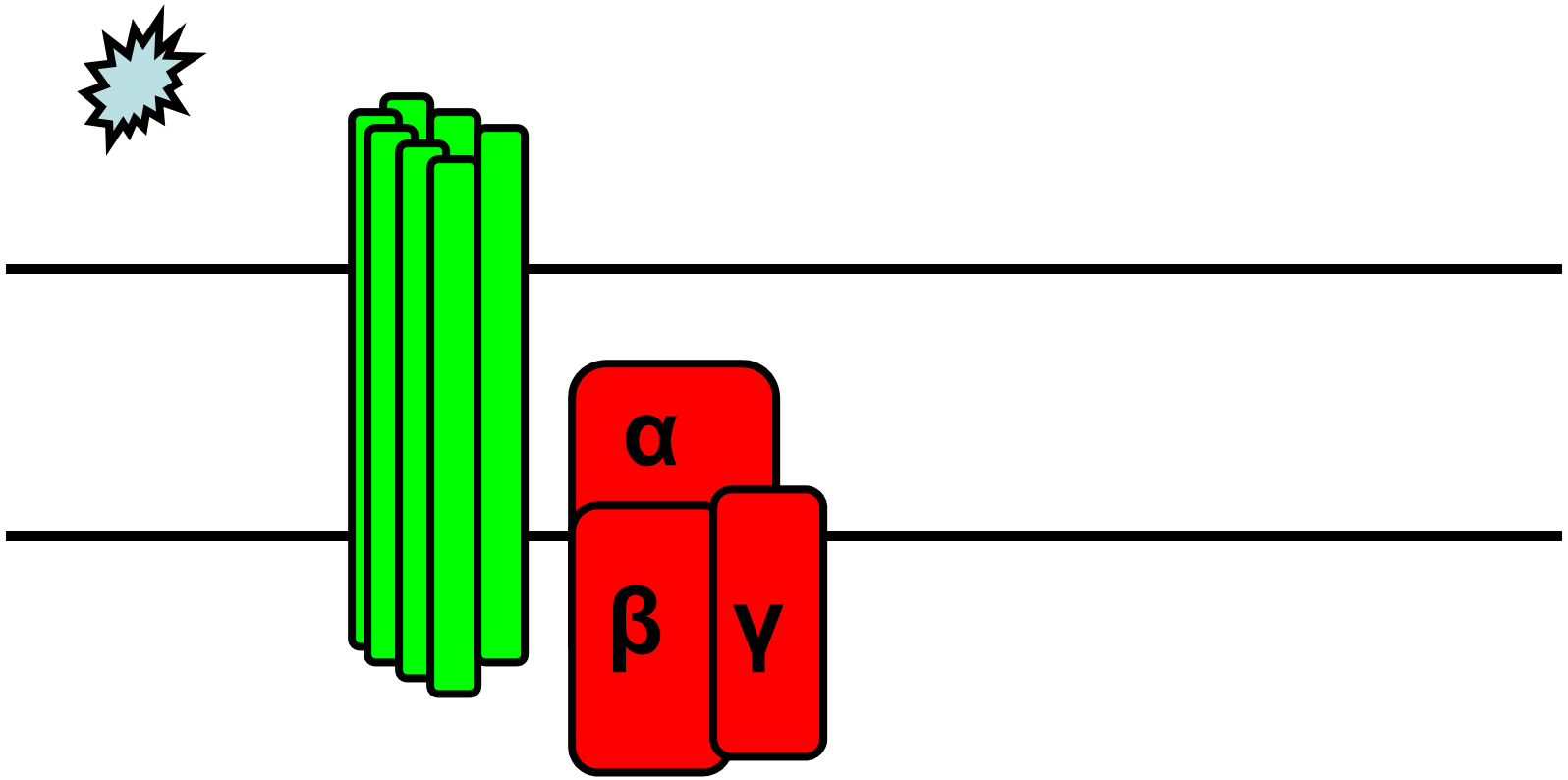
Receptors

- **Types**
  - Ion channels
  - G-protein coupled receptors
  - Receptor-enzymes
  - Cytosolic-nuclear receptors
- **Act as transducer proteins**
  - Receptor-effector signal transduction
  - Post-receptor signal transduction provides for amplification of the signal

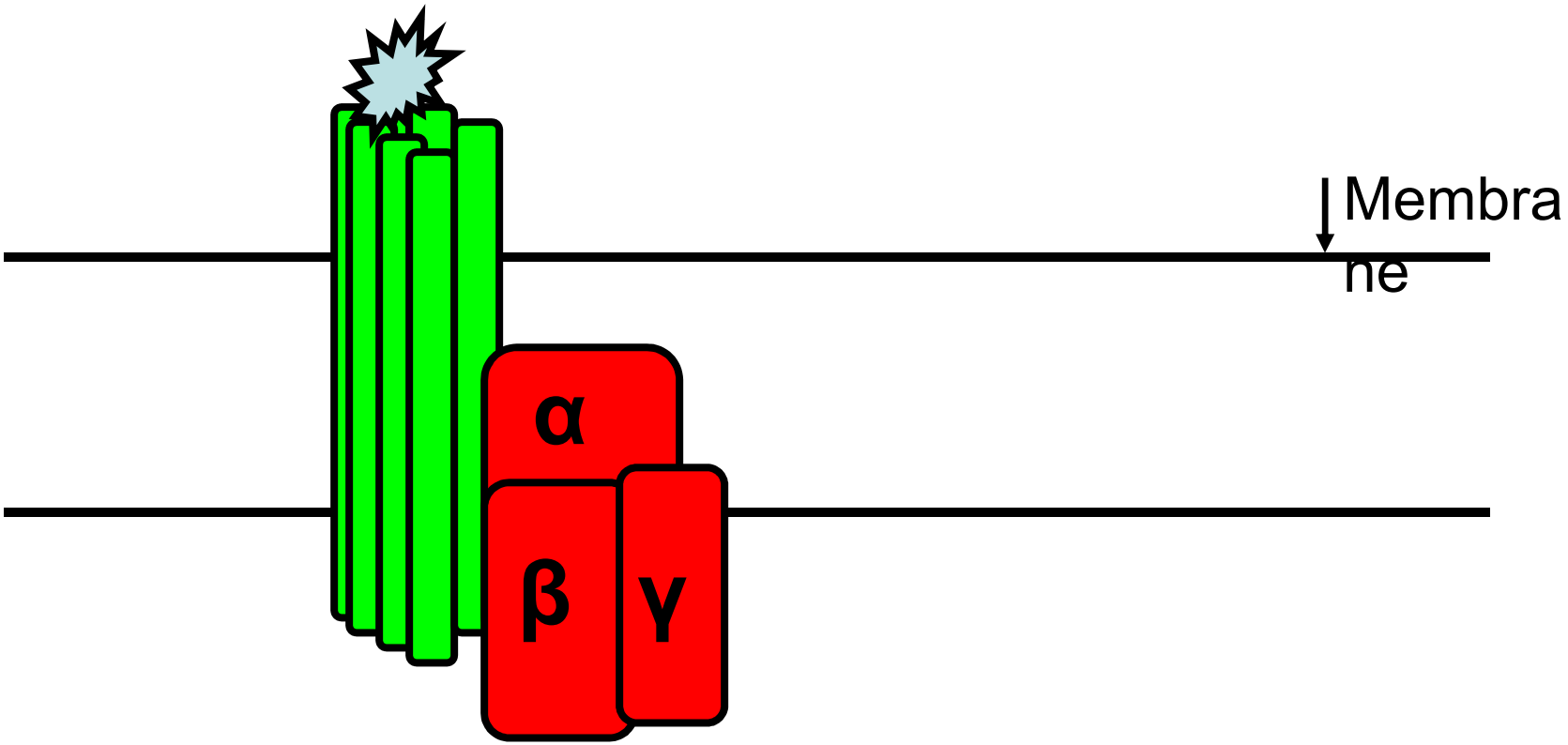
Ligand-gated Ion Channels



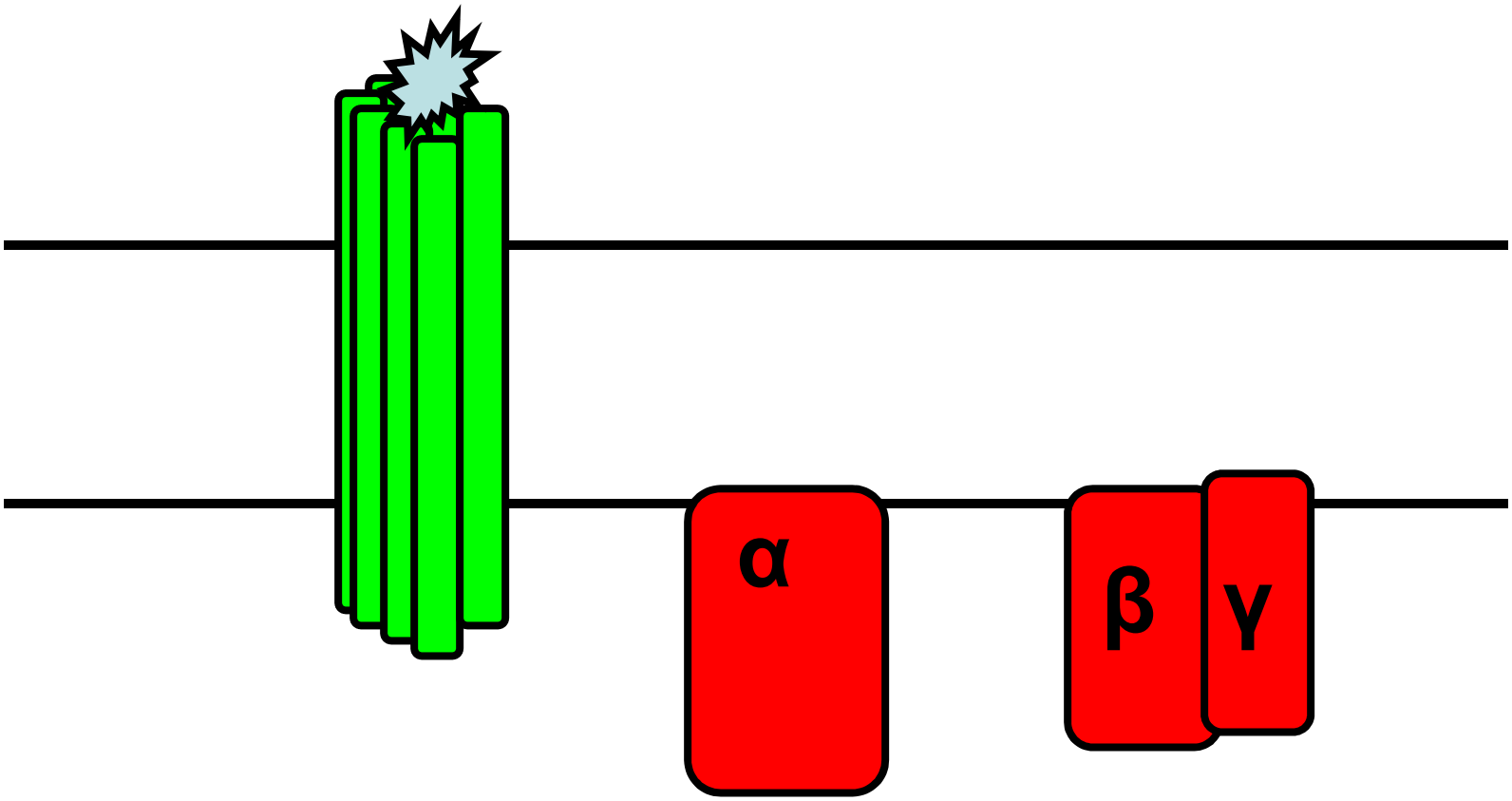
G-protein coupled receptors



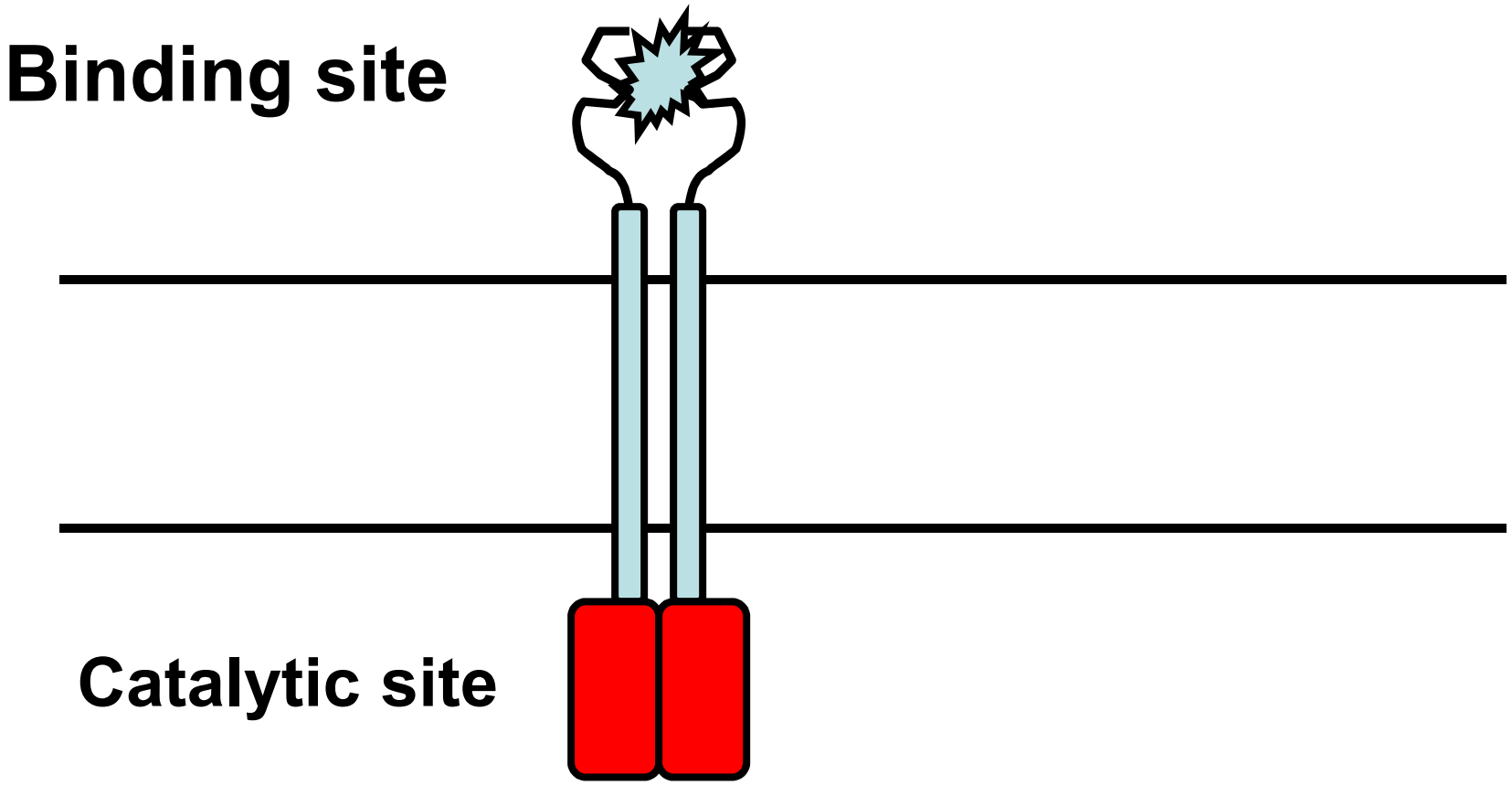
G-protein coupled receptors



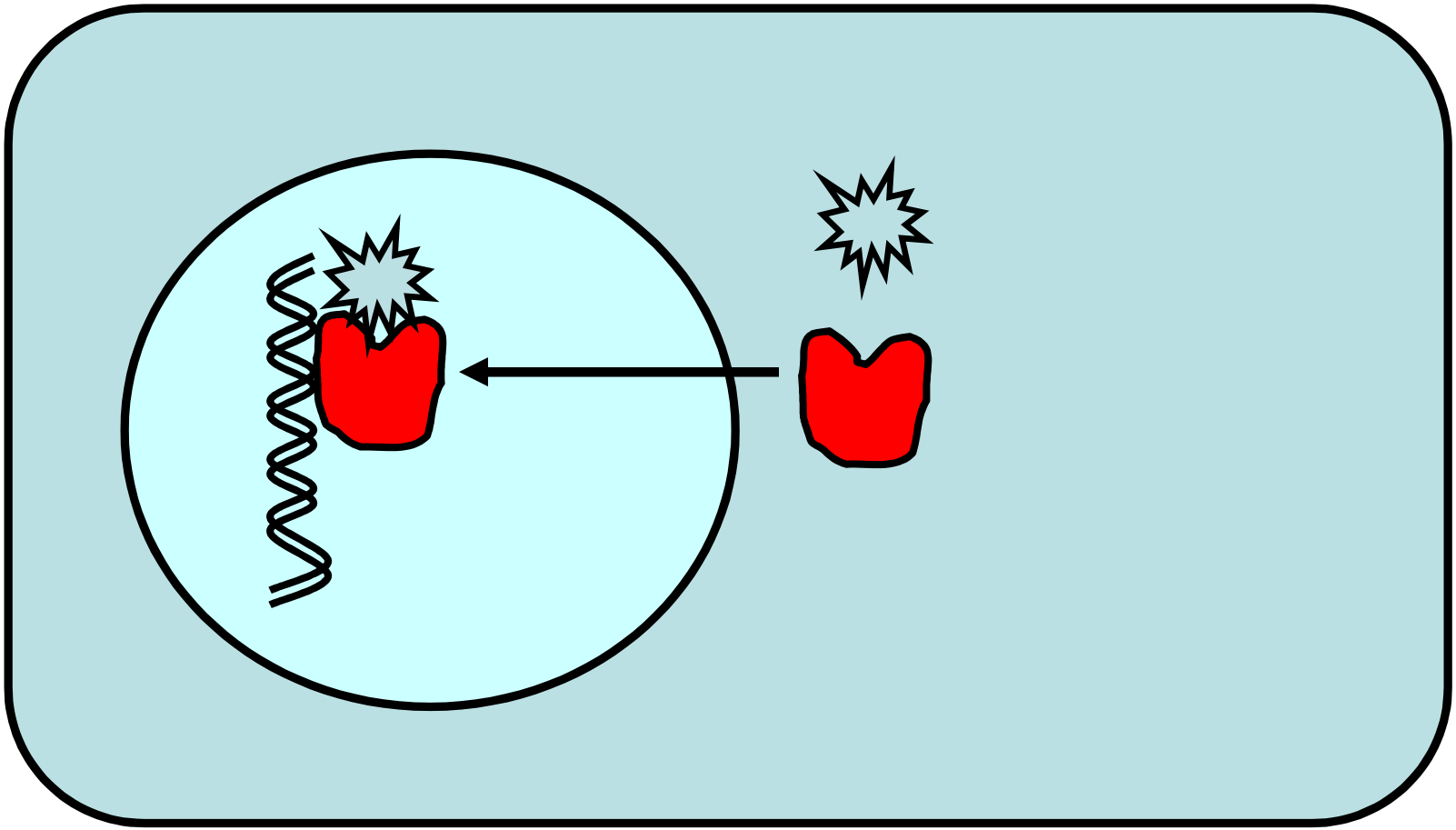
G-protein coupled receptors

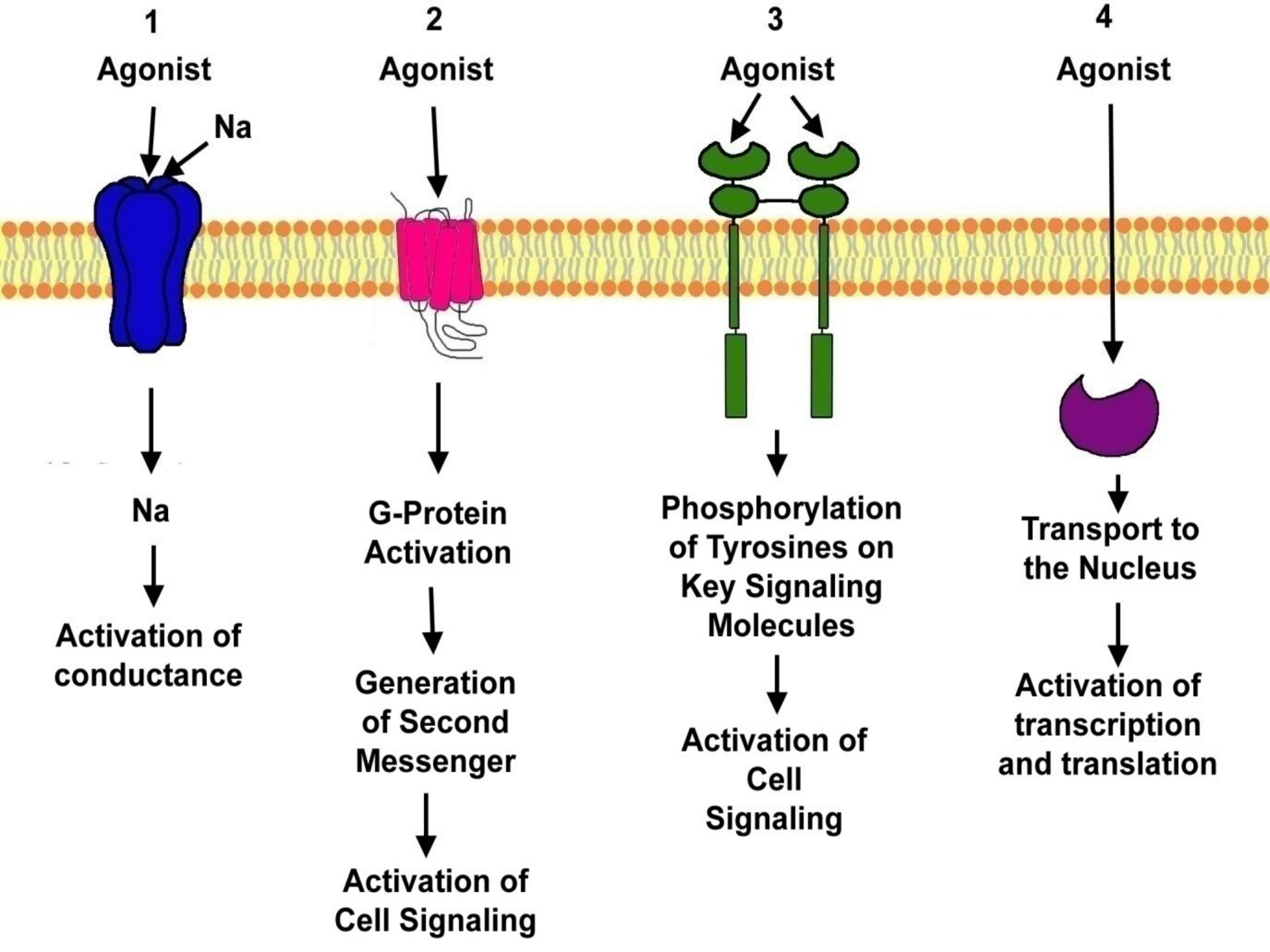


Receptor-enzyme



Cytosolic-Nuclear receptors




1
Agonist
Na
Na
Activation of conductance
2
Agonist
G-Protein Activation
Generation of Second Messenger
Activation of Cell Signaling
3
Agonist
Phosphorylation of Tyrosines on Key Signaling Molecules
Activation of Cell Signaling
4
Agonist
Transport to the Nucleus
Activation of transcription and translation

HOW DRUGS ACT

l. Enzyme Inhibition:

Drugs act within the cell by modifying normal biochemical reactions.
Enzyme inhibition may be reversible or non reversible;
competitive or non-competitive.
Antimetabolites may be used which mimic natural metabolites. Gene functions may be suppressed.

2. Drug-Receptor Interaction:

Drugs act on the cell membrane by physical and/or chemical interactions-
usually through specific drug receptor sites known to be located on the membrane.
Some receptor sites have been identified with specific parts of proteins and nucleic acids. In most cases, the chemical nature of the receptor site remains obscure.

3. Non-specific Interactions:

Drugs act exclusively by physical means outside of cells. These sites include external surfaces of skin and gastro-intestinal tract. Drugs also act outside of cell membranes by chemical interactions. Neutralization of stomach acid by antacids is a good example

Factors Governing Drug Action

Two factors that determine the effect of a drug on physiologic processes are

1. Affinity

2.Intrinsic activity.

Affinity is a measure of the tightness that a drug binds to the receptor.

Intrinsic activity is a measure of the ability of a drug once bound to the receptor to generate an effect activating stimulus and producing a change in cellular activity.

Classification of a drug based on drug-receptor interactions:

- **Agonist:** Drug that binds to receptors and initiates a cellular response; has affinity and efficacy. Agonists promote the active state.

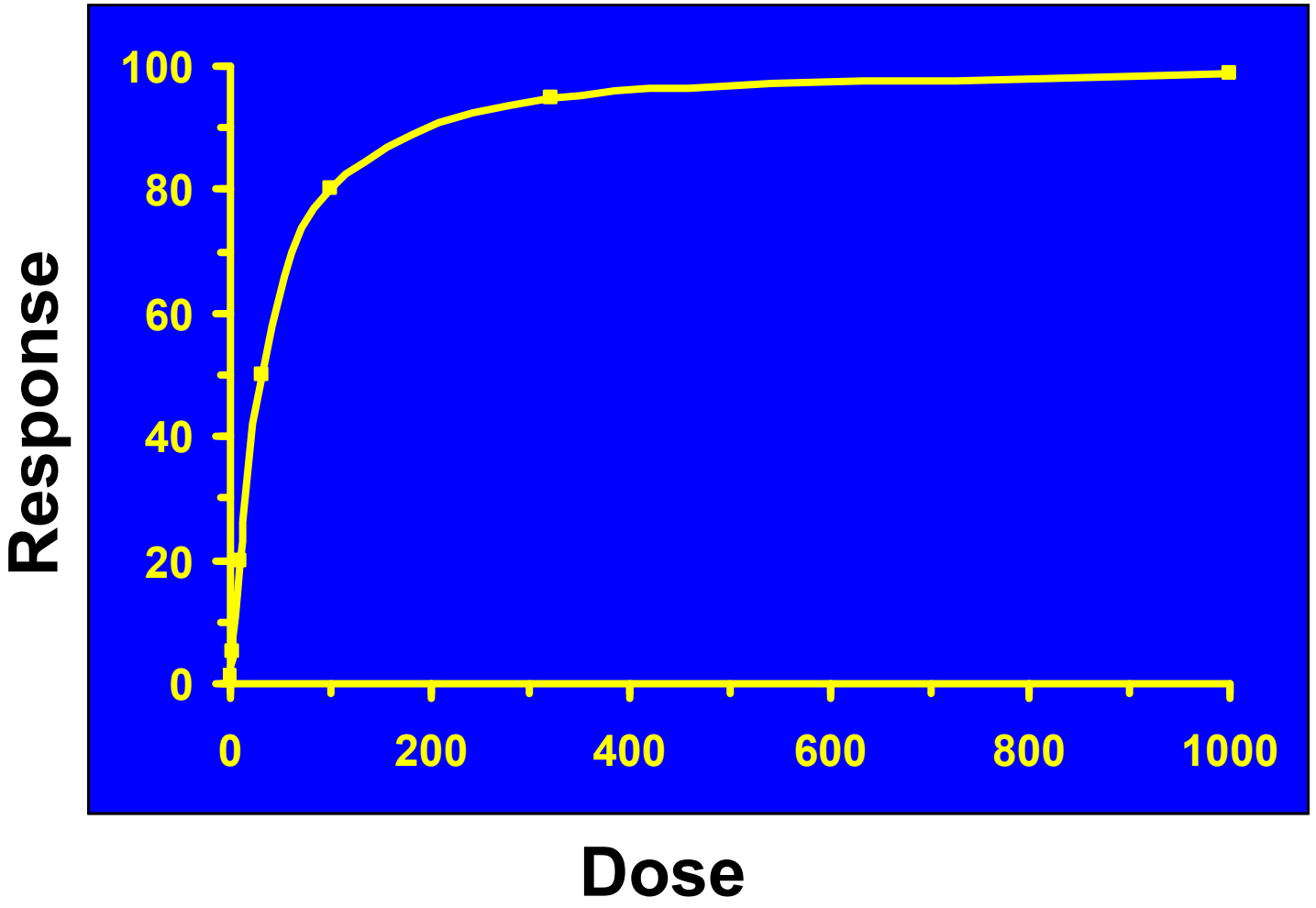
- **Antagonist:** drug that binds to receptors but cannot initiate a cellular response, but prevent agonists from producing a response; affinity, but no efficacy. Antagonists maintain the active-inactive equilibrium.

cont.

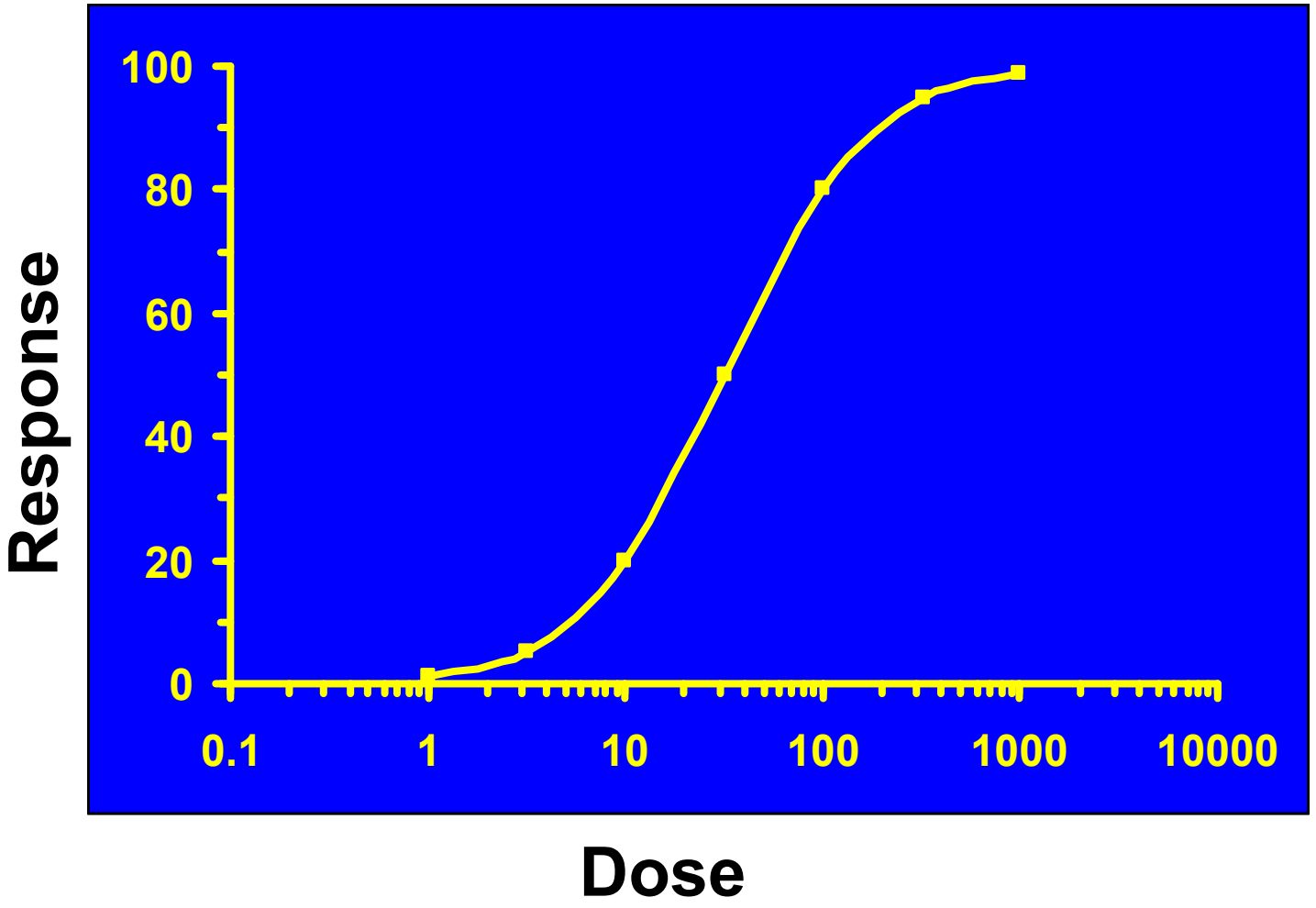
- **Partial agonists:** Drug that, no matter how high the dose, cannot produce a full response.

- **Inverse agonist:** Drug that binds to a receptor to produce an effect opposite that of an agonist. Stabilizes receptors in the inactive state.

Dose-response curve



Dose-response curve

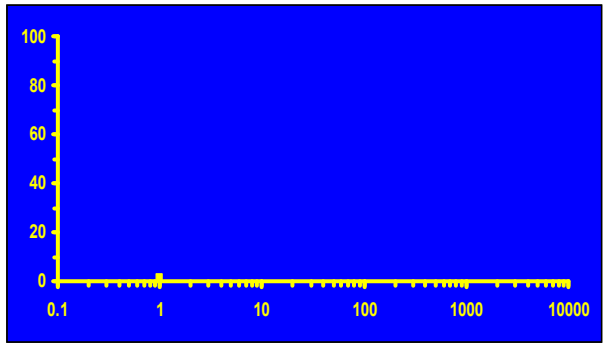


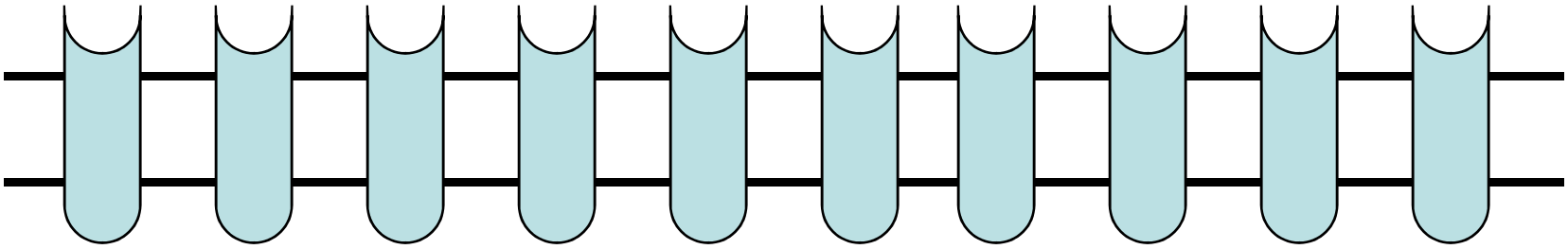
Concepts to remember:

Threshold: Dose that produces a just-noticeable effect.

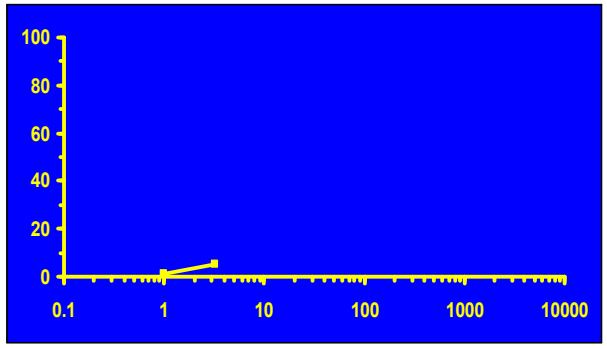
ED_{50}: Dose that produces a 50% of maximum response.

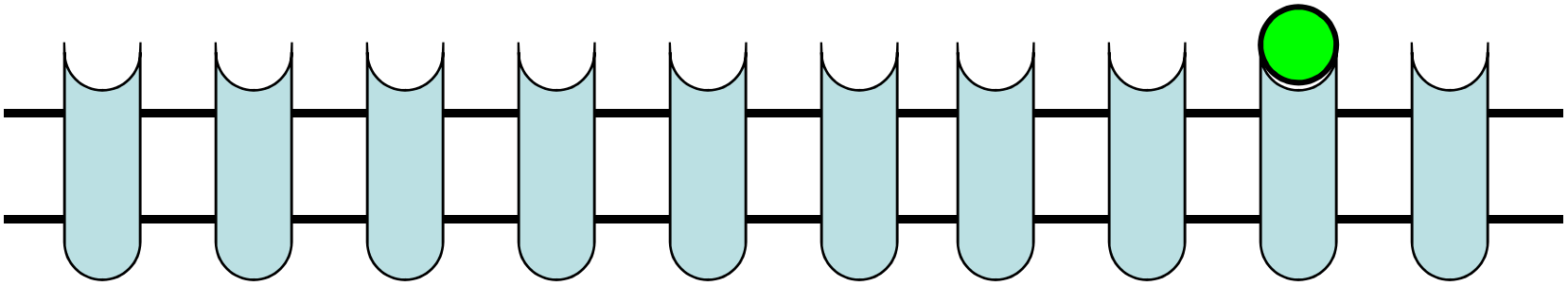
Ceiling: Lowest dose that produces a maximal effect.



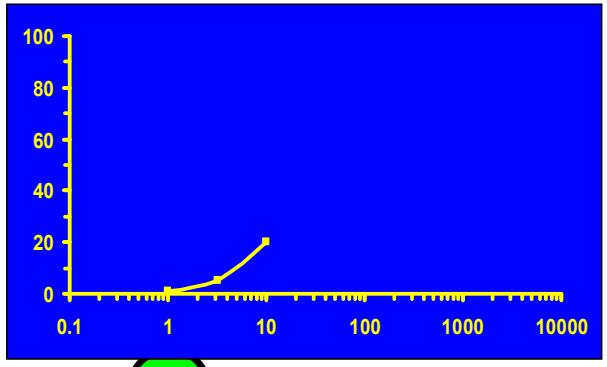


= Agonist

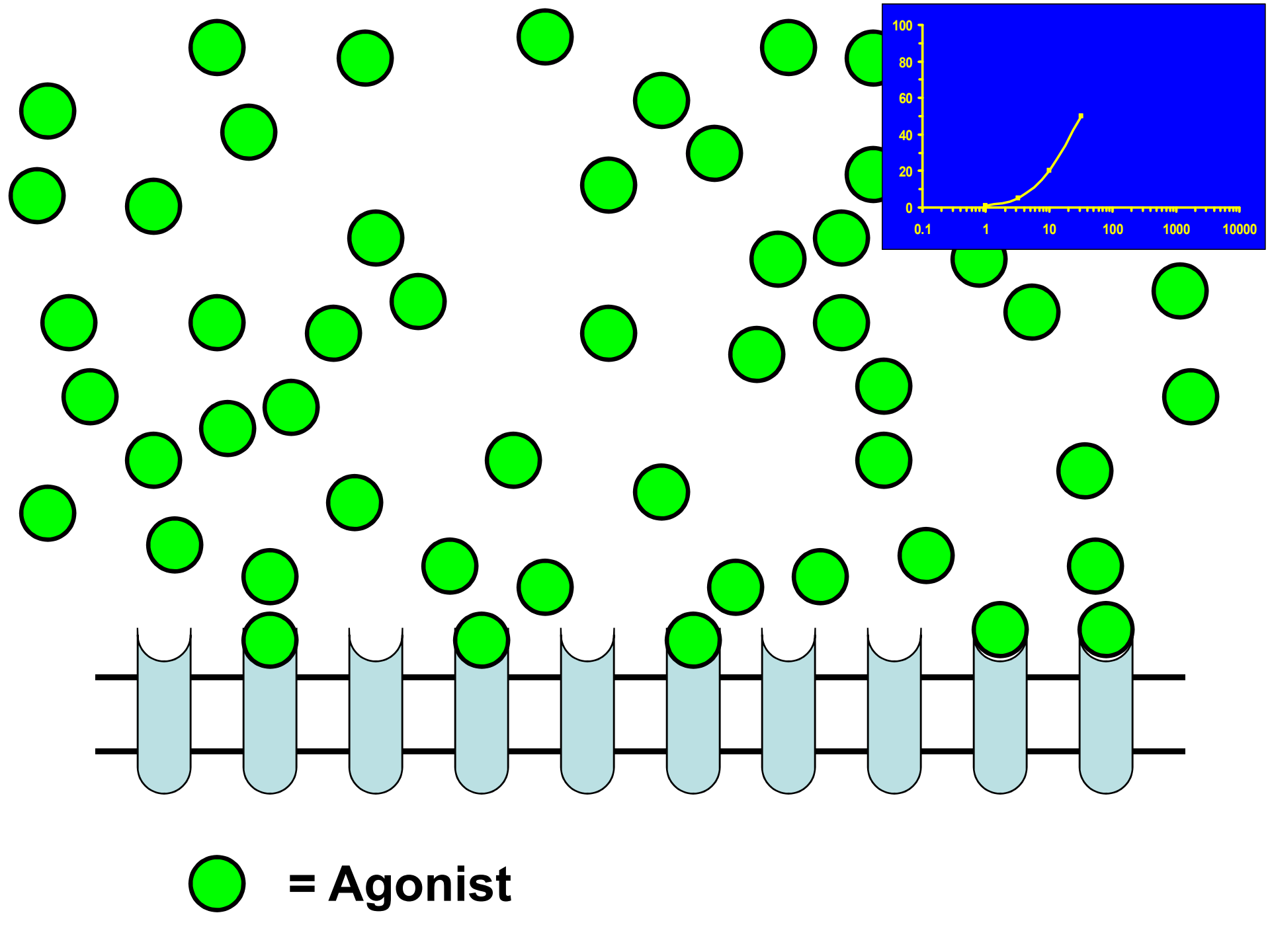


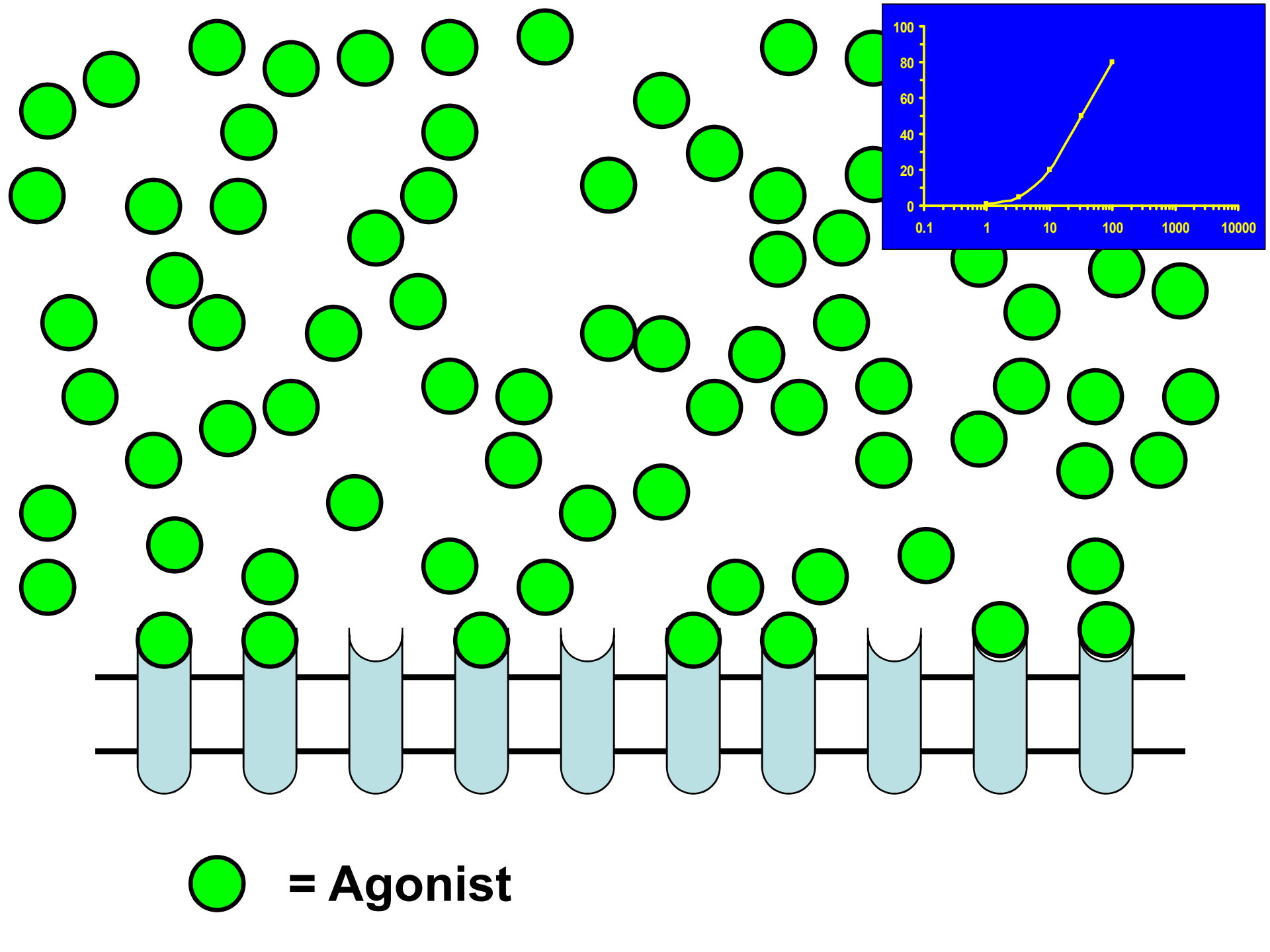


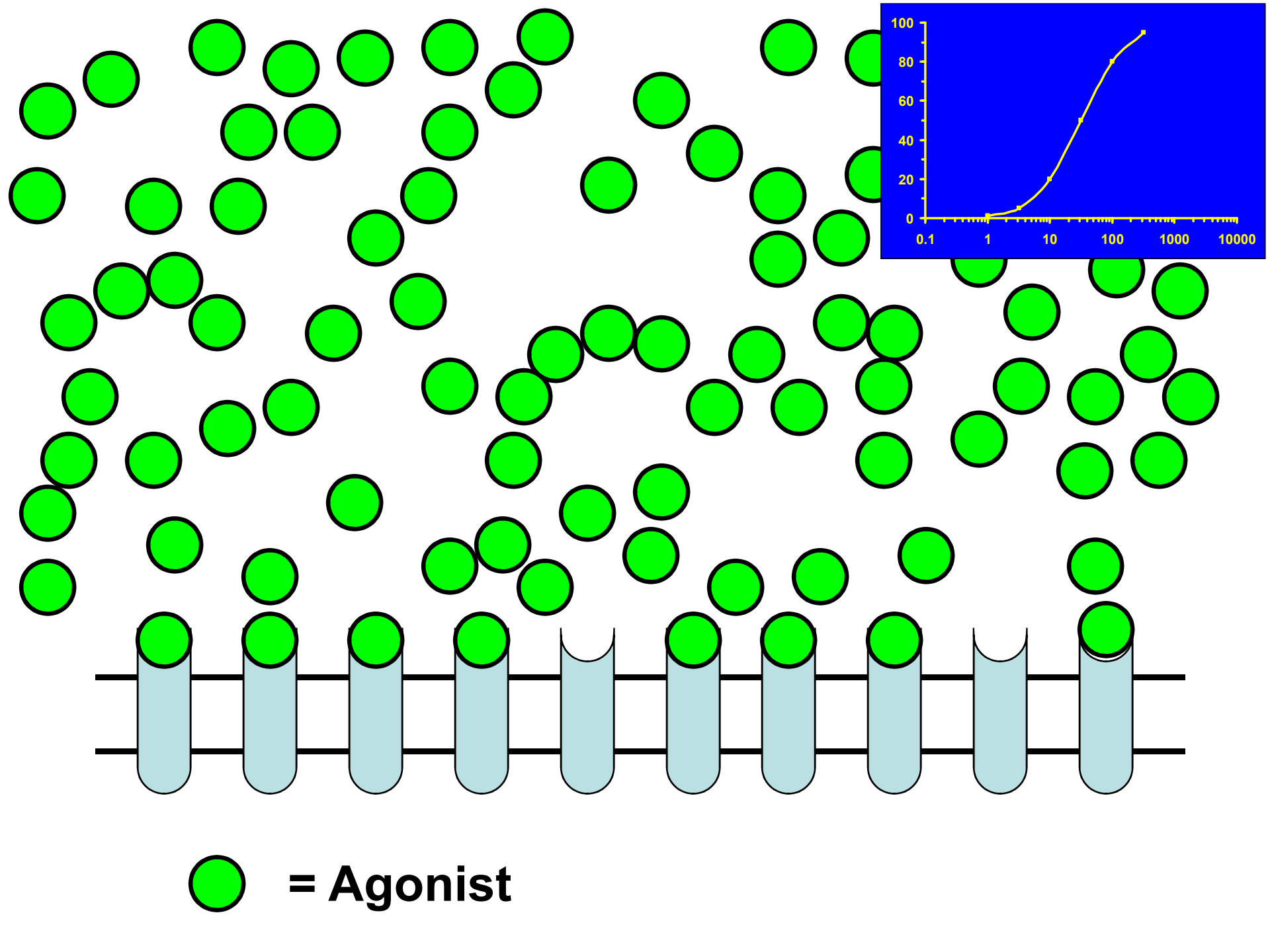
= Agonist

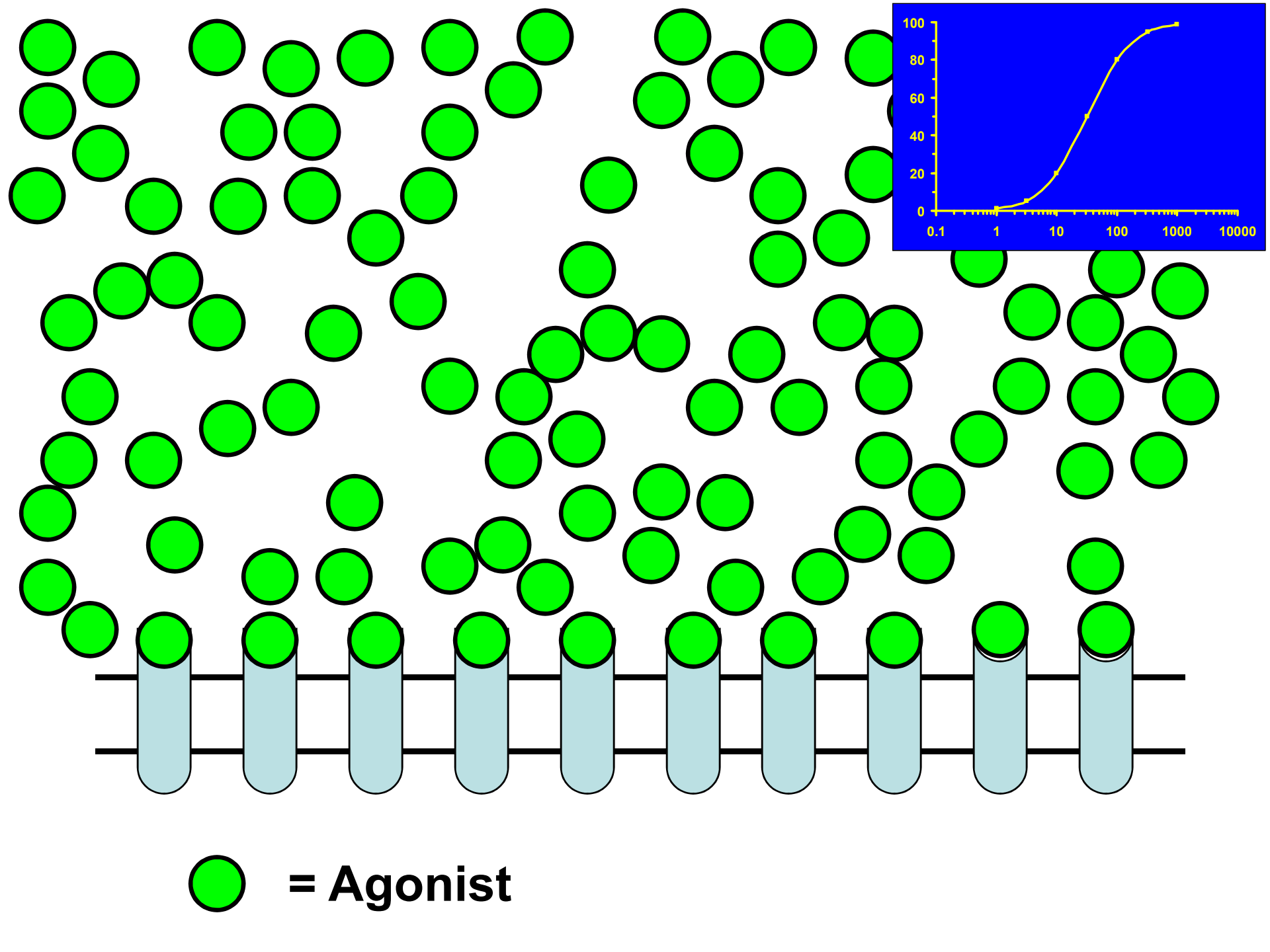


= Agonist

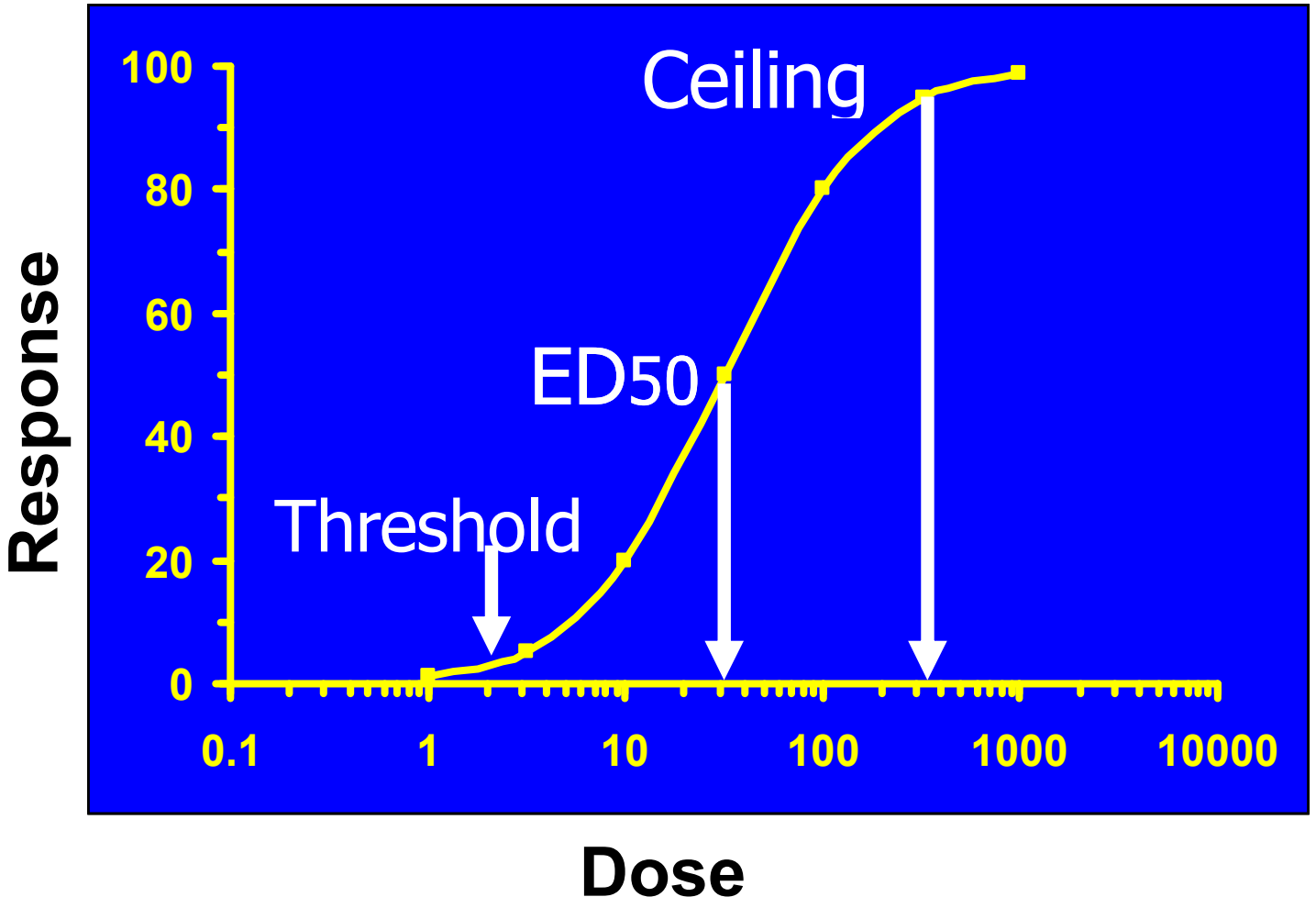

100
80
60
40
20
0
0.1
1
10
100
1000
10000
= Agonist


100
80
60
40
20
0
0.1
1
10
100
1000
10000
= Agonist

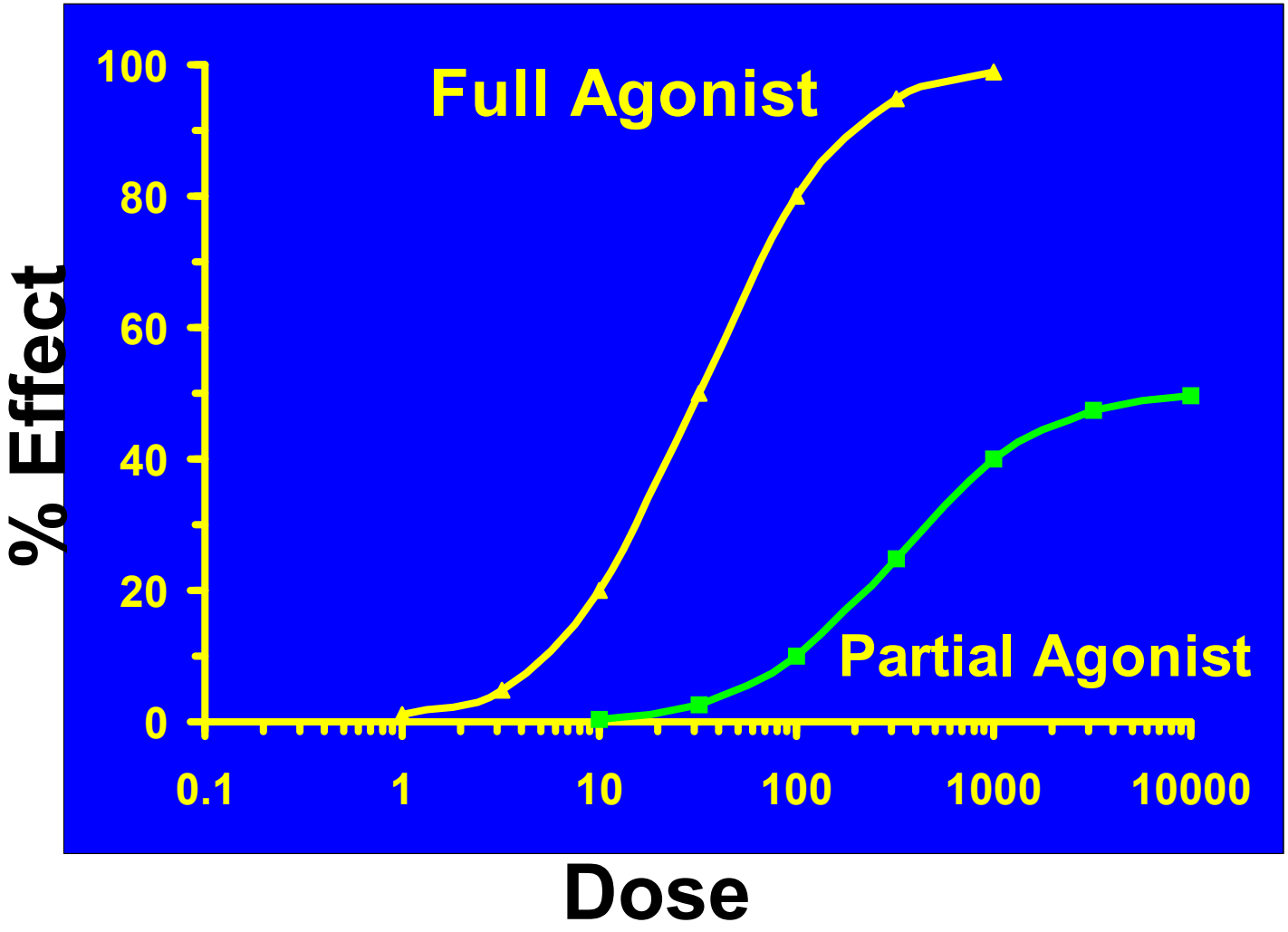

100
80
60
40
20
0
0.1
1
10
100
1000
10000
= Agonist


100
80
60
40
20
0
0.1
1
10
100
1000
10000
= Agonist

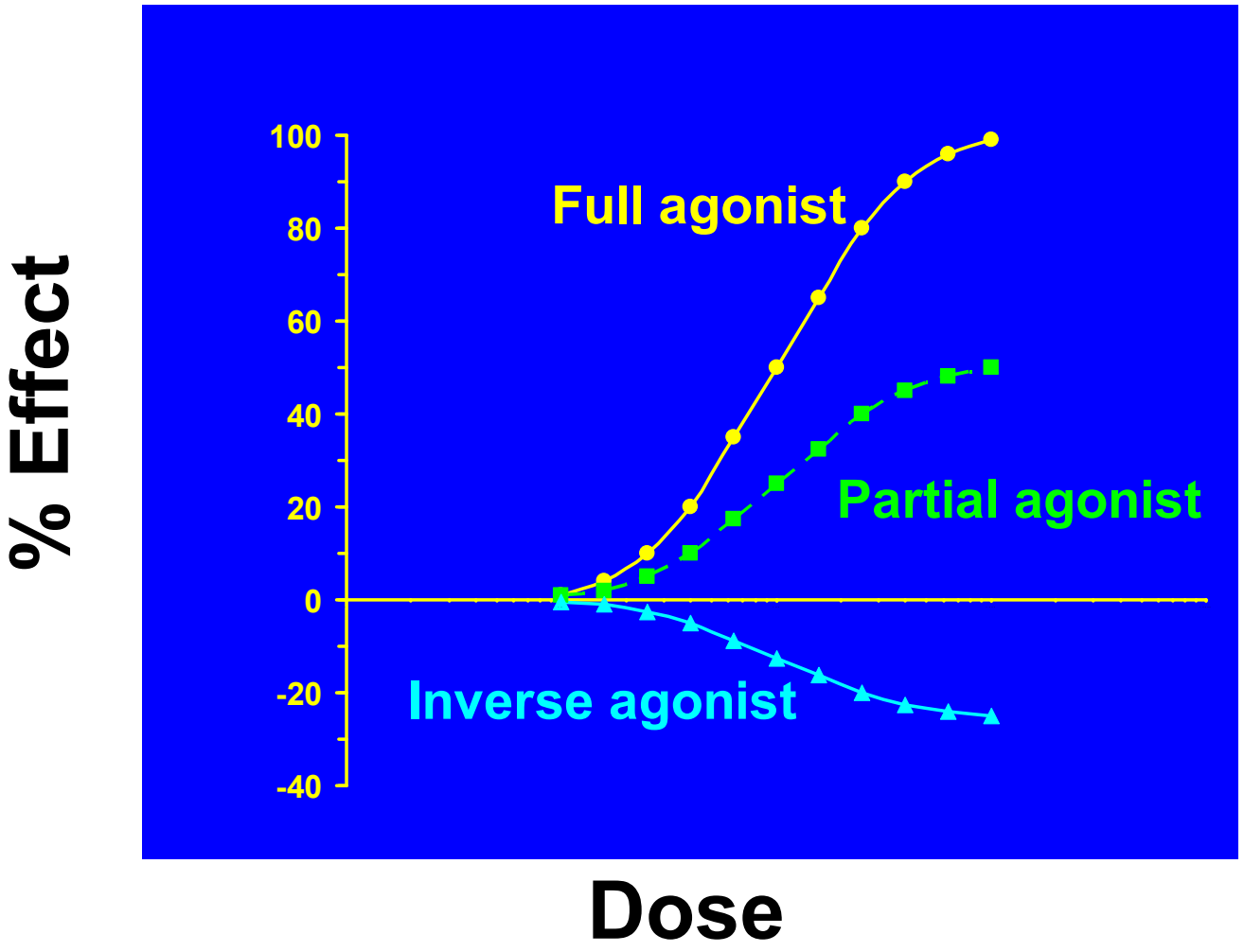
Dose-response curve



Full vs Partial agonists



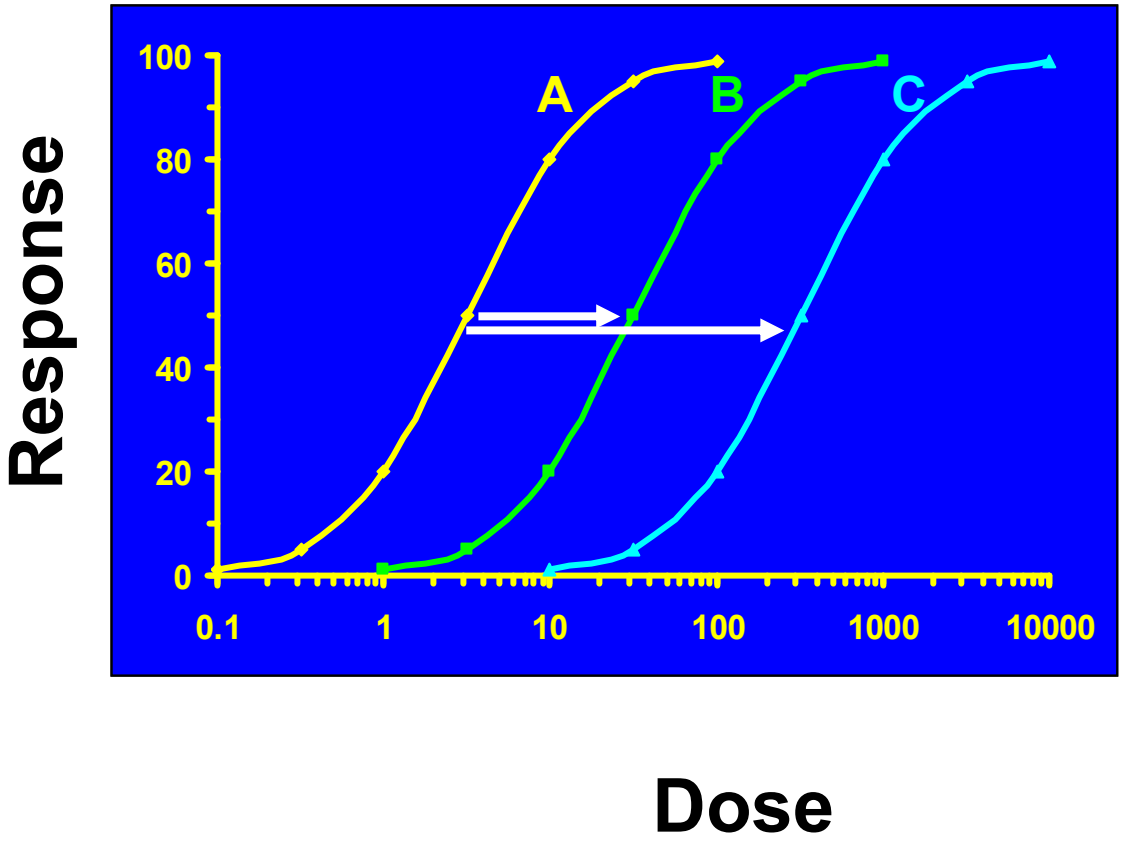
Inverse Agonist



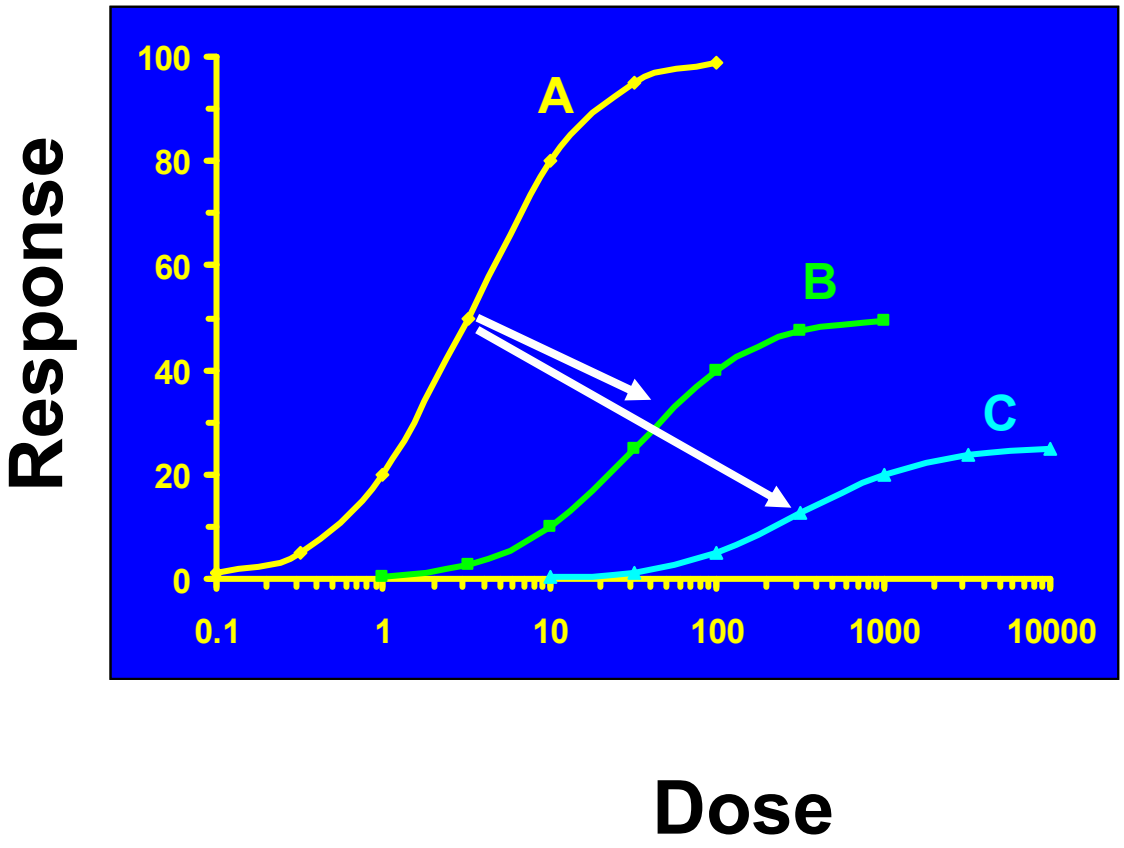
Antagonists

- **Competitive:** Antagonist binds to same site as agonist in a reversible manner.
- **Noncompetitive:** Antagonist binds to the same site as agonist irreversibly.
- **Allosteric:** Antagonist and agonist bind to different site on same receptor
- **Physiologic:** Two drugs have opposite effects through differing mechanisms

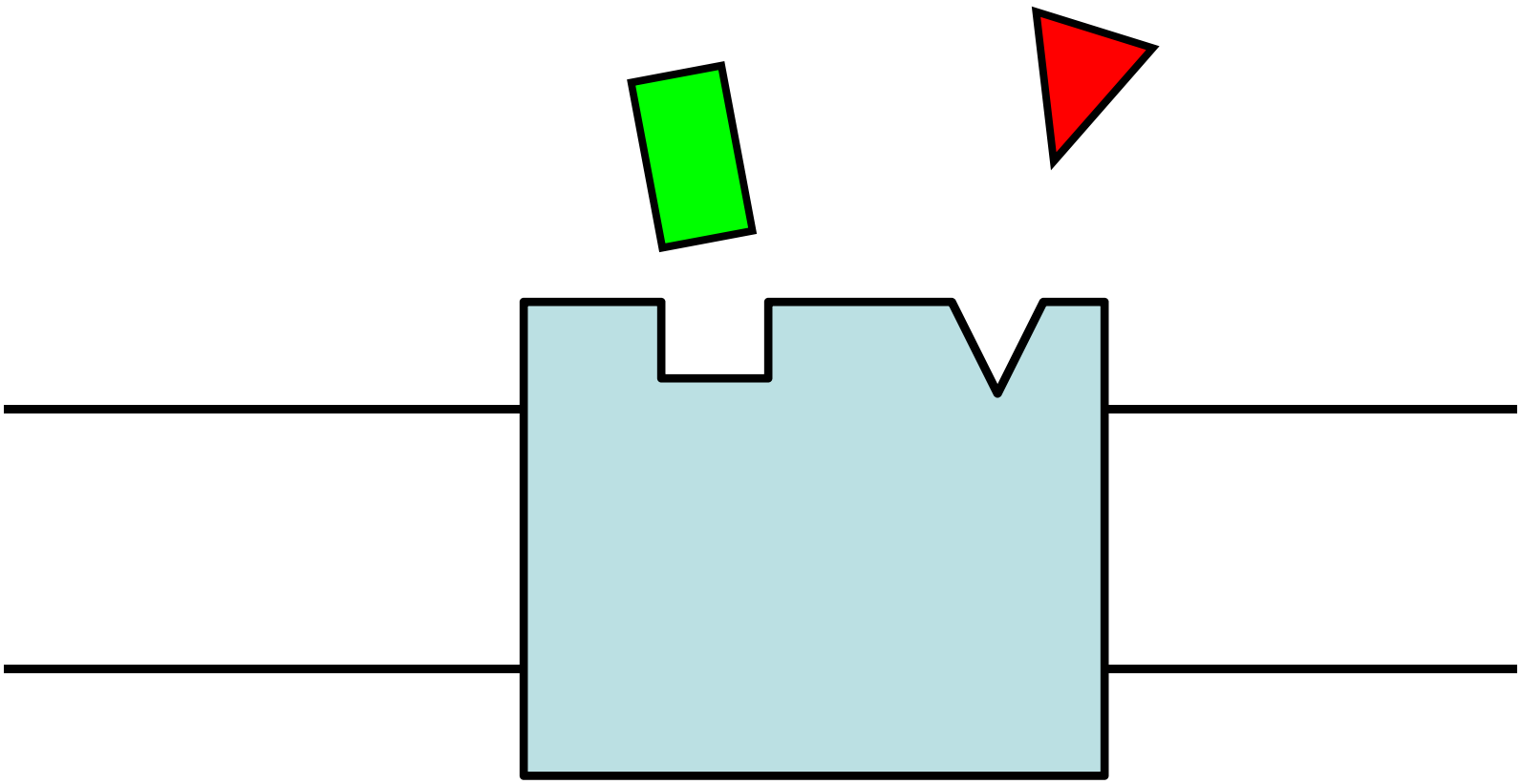
Competitive antagonists



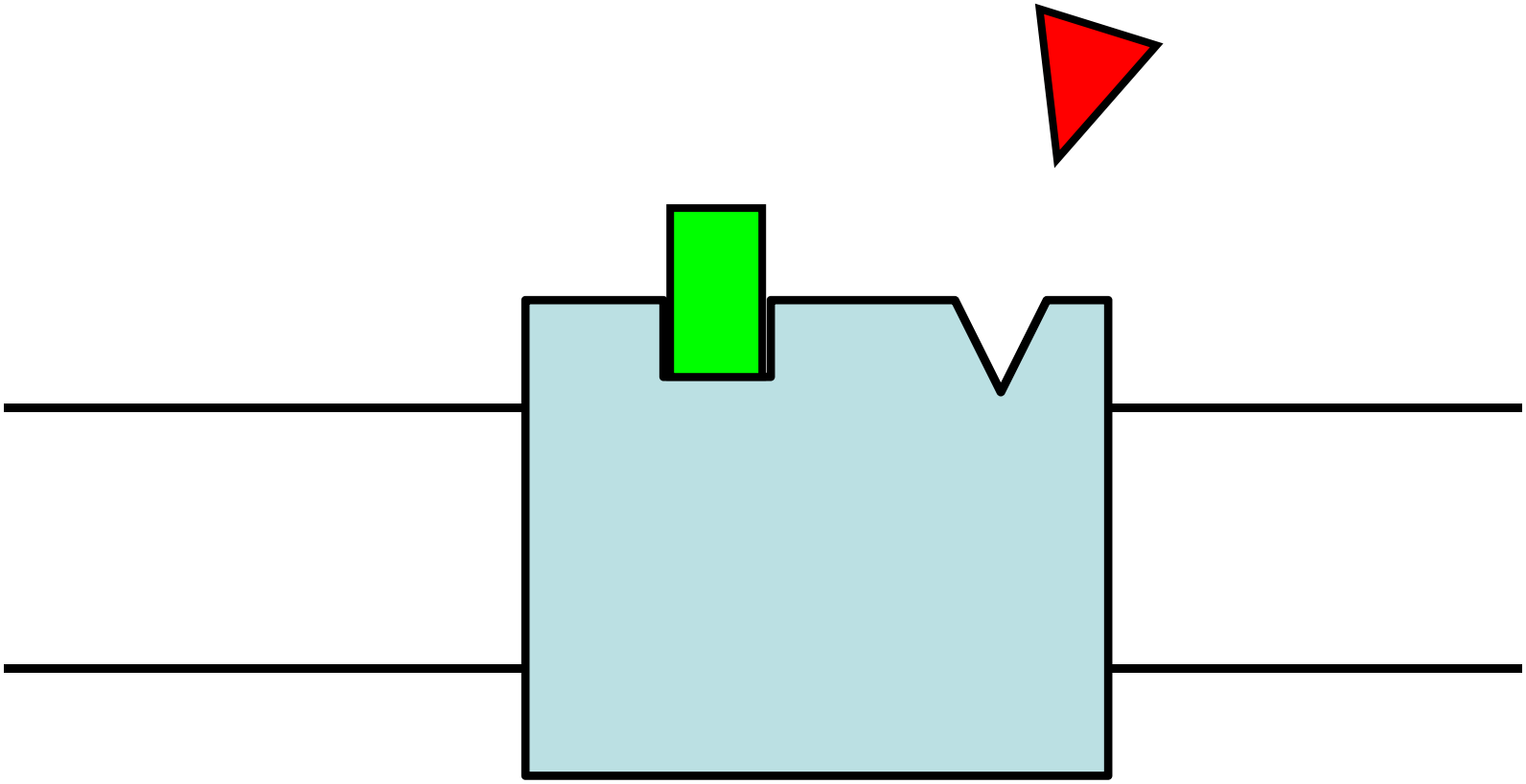
Noncompetitive antagonists



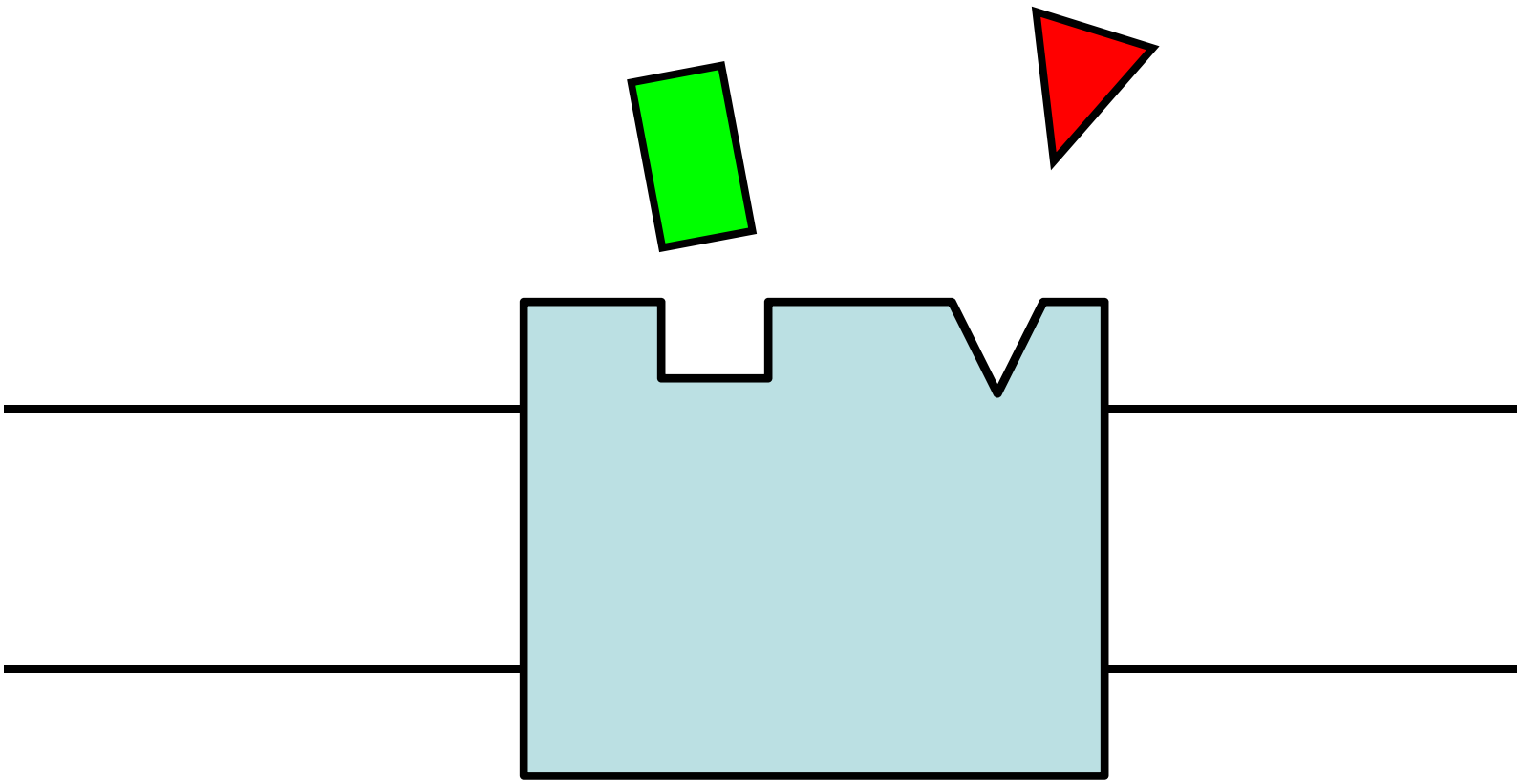
Allosteric Antagonism



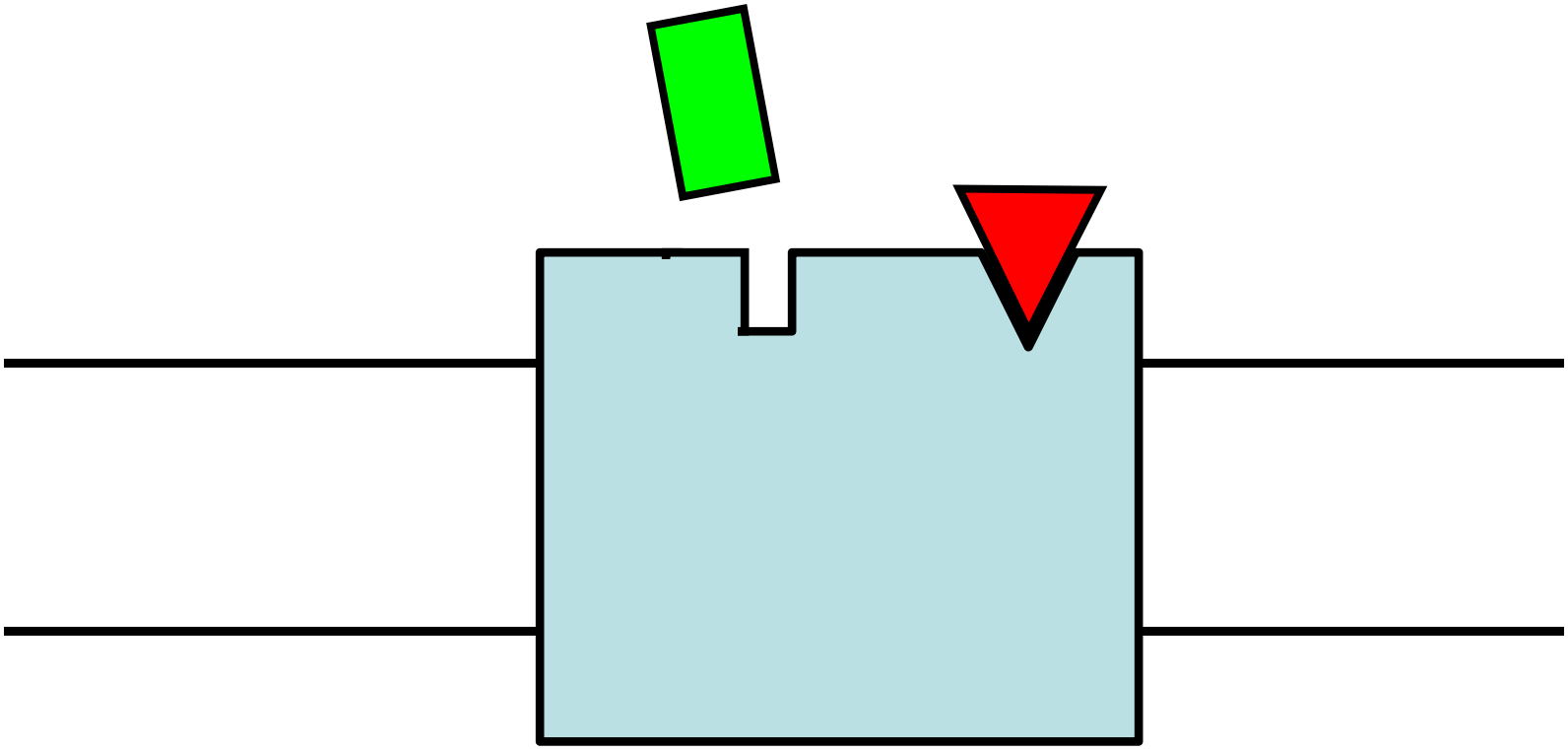
Allosteric Antagonism



Allosteric Antagonism



Allosteric Antagonism



Prodrug

- A **prodrug** is a medication that is administered in a pharmacologically inactive form which is then converted to an active form through a normal metabolic process
- A prodrug is a precursor chemical compound of a drug
- Instead of administering a drug, a prodrug might be used instead to improve how a medicine is absorbed, distributed, metabolized, and excreted (ADME)

Types of Prodrug

- Type 1
  - bioactivated inside the cells
- Type2
  - bioactivated outside the cells

Hard drugs

- *Hard drugs* are drugs that lead to physical addiction.
- Many countries do not allow people to make, sell or use some of them, other than for medical purposes.
- Examples of such drugs are heroin, methamphetamine (meth), cocaine, alcohol and nicotine.

Soft drugs

- *Soft drugs* are not thought to cause physical addiction.
- Examples of soft drugs are mescaline, psilocybin etc.
- While they do not cause physical addiction, some of them may still lead to psychological dependence

ADME/T properties of drug

- Adsorption
- Distribution
- Metabolism
- Excretion
- Toxicity

Lipinski rule

- Molecular weight - <= 500
- Log P value - <= 5
- No. of Donors - <= 5
- No. of Acceptors - <=10

Drug metabolism

- The body typically deals with a foreign compound by making it more water-soluble, to increase the rate of its excretion through the urine.
- There are many different processes that can occur; the pathways of drug metabolism can be divided into:
  - Phase I
  - Phase II
- Drugs can undergo one of four potential biotransformations:
  - Active Drug to Inactive Metabolite
  - Active Drug to Active Metabolite
  - Inactive Drug to Active Metabolite
  - Active Drug to Toxic Metabolite (Biotoxification)

Phase I

- Includes oxidative, reductive, and hydrolytic reactions.
- In these type of reactions, a polar group is either introduced or unmasked, so the drug molecule becomes more water-soluble and can be excreted.
- Reactions are non-synthetic in nature and in general produce a more water-soluble and less active metabolites.
- The majority of metabolites are generated by a common hydroxylating enzyme system known as Cytochrome P450.

Phase II

- These reactions involve covalent attachment of small polar endogenous molecule such as glucuronic acid, sulfate, or glycine to form water-soluble compounds.
- This is also known as a *conjugation reaction*.
- The final compounds have a larger molecular weigh

Thanks